



UNIVERSIDAD PANAMERICANA

Facultad de Derecho

Posgrado en Derecho

Con Reconocimiento de Validez Oficial ante la Secretaría de Educación Pública, bajo acuerdo número 985162 de fecha 17 de agosto de 1998.

Análisis de la aplicación de las categorías de responsabilidad jurídica a casos de uso de sistemas autónomos de Inteligencia Artificial para la toma de decisiones y la utilización de la Ética empresarial. Estudio del caso *Watson Health* de IBM

Tesis que para obtener el grado de

Doctor en Derecho

Sustenta el alumno

Mtro. José Luis Hernández Sánchez

Director de la Tesis

Dr. Jorge Cerdio Herrán

Ciudad de México, agosto de 2022

Dedicatoria

A Dios por la oportunidad de vivir y estar rodeado de personas excepcionales.

A mi madre y hermano, ejemplos lucha, inteligencia y amor perenne.

A mi padre †

A mi esposa y a mi hijo, razones suficientes para luchar hasta el último respiro

A Magdalena Sánchez Medina y a Jesús Anaya Gutiérrez †

A Jorge Cerdio Herrán, imposible alcanzar esto sin tu sensibilidad e inteligencia. El camino de conocimientos compartidos hoy rinde frutos

A José Alberto Ross, por acercar a las generaciones del Doctorado en Derecho a la filosofía de Carlos Llano Cifuentes

A Diego Valadés Ríos por acompañarme en la construcción de mi formación académica previa. La distancia no es olvido, es recuerdo y presente

A Rafael Estrada Michél, por tenderme la mano en los momentos difíciles.
¡Maestro y Amigo Dilecto!

A la Facultad de Derecho de la Universidad Panamericana quien me abrió las puertas para cerrar este ciclo de formación; y, con especial aprecio a los integrantes del sínodo doctoral. Gracias por leer esta investigación

Agradecimientos

Mi agradecimiento a las personas que durante estos últimos años de mi vida han estado presentes en este esfuerzo y contribuido de manera diversa a su conclusión. En particular a:

Aline Guerrero	Karen González
Aleida Maldonado	Lorena Leyva
Alejandro Carlos	Marco Antonio Gómez
Ángel González	Manuel Gracián
Aníbal Monasterio	Mayté Calcáneo
Arturo Garita	Michelle Azuara
Atocha Aliseda	Miguel Sarre
Carlos Esquivel	Rafael Martínez
Claudia Elena Pérez	Ricardo Alcántar
Eduardo Almanza	Rodolfo Gómez
Enrique Siqueiros	Roberto Gallo
Fátima Ríos	Salvador Cárdenas
Francisco Guevara	Salvador López
Gerardo Mojardín	Sandra Boy
Germán Castillo	Valentín Vargas
Hugo Ortiz	Yesica Espinosa
Isabel Cornejo	
Isabel Mojardín	
Ivette Uribe	
Jaime García	
Jaime López	
Jesús Jiménez	
José Luis Mora	
Julio Tellez	
Jesús Esquivel	

- Y la Madre le preguntó al Gerente-: *Cada Amigo Artificial es único, ¿verdad?*

-Así es, señora.

- ¿Y qué es lo que hace única a esta... Klara?

- *Klara posee tantas cualidades únicas que no podríamos pasar aquí la mañana repasándolas. Pero si tuviera que destacar una, bueno, diría que son sus ganas de observar y aprender. Su habilidad para absorber y relacionar todo lo que ve a su alrededor es asombrosa. Como resultado, ahora mismo posee un entendimiento más sofisticado que cualquier AA de esta tienda...*

Kazuo Ishiguro

Klara y el Sol

Editorial Anagrama, 2021

Contenido

	Pág.
<u>Introducción</u>	1
<u>Capítulo I. Inteligencia Artificial</u>	28
1) <u>Inteligencia Artificial basada en Sistemas Expertos Decisionales</u>	36
A) <u>Evolución y conceptos cardinales</u>	38
B) <u>Sistemas Expertos Decisionales</u>	50
2) <u>Inteligencia Artificial basada en Machine Learning y Deep Learning</u>	81
A) <u>Machine Learning</u>	89
B) <u>Deep Learning</u>	106
<u>Capítulo II. Inteligencia Artificial basada en Machine Learning y Deep Learning. Problemas de funcionamiento y rendición de cuentas</u>	136
1) <u>Problemas de funcionamiento</u>	140
A) <u>Datos y sesgos</u>	142
B) <u>De la opacidad a la interpretabilidad y explicabilidad algorítmicas</u>	157
2) <u>Rendición de cuentas</u>	176
A) <u>Transparencia y privacidad</u>	177
B) <u>Auditabilidad algorítmica</u>	208

<u>Capítulo III. IBM Watson y el caso <i>Watson Health</i></u>	236
1) <u><i>IBM Watson</i></u>	242
A) <u>Características y funciones</u>	244
B) <u>Casos y evaluación del sistema</u>	266
2) <u><i>El caso Watson Health</i></u>	289
A) <u><i>Healthcare Technology</i></u>	311
B) <u><i>Watson Health: características y problemática de su uso</i></u>	327
 <u>Capítulo IV. Categorías de la responsabilidad jurídica, su aplicación a casos de sistemas de Inteligencia Artificial para la toma de decisiones y el uso de la Ética empresarial</u>	363
1) <u><i>Las tres dimensiones de la responsabilidad</i></u>	367
A) <u>Categorías de la responsabilidad jurídica</u>	386
B) <u>Categorías de la responsabilidad algorítmica y otros conceptos</u>	429
C) <u>Categorías de la responsabilidad moral</u>	460
2) <u><i>El caso Watson Health y la aplicación de las tres dimensiones de la responsabilidad</i></u>	488
A) <u>Metodología para la medición de riesgos legales, éticos y tecnológicos</u>	488
B) <u>Problematización del caso <i>Watson Health</i> de IBM</u>	505
<u>Reflexiones finales</u>	527

<u>Glosario</u>	533
<u>Índice de tablas, gráficas y diagramas</u>	546
<u>Fuentes</u>	553
<u>Anexos</u>	609

Introducción

¿Las máquinas piensan? ¿aprenden? ¿toman decisiones por sí mismas? ¿resuelven problemas como los seres humanos? ¿tienen conciencia? ¿corrigen sus errores por sí mismas? Si a todo esto dijéramos que sí y estas máquinas tomaran una decisión errónea, no la más acertada o equivocada, sin consentimiento o la participación del ser humano, quién o quiénes serían los sujetos de la responsabilidad jurídica o, quizás antes debamos preguntarnos si el sistema jurídico contiene la categoría de responsabilidad adecuada para resolver este tipo de casos y saber si se puede trasladar la responsabilidad a las máquinas o liberar de responsabilidad a los seres humanos. También, deberíamos pensar si el sistema de justicia y sus estándares probatorios son compatibles con esta tecnología y permiten aplicar sus reglas como la equidad probatoria o el principio de la carga de la prueba, sin que el ser humano más cercano a la tecnología cargue con toda la responsabilidad y reciba una sentencia condenatoria eximiendo de toda responsabilidad a la empresa o al sistema de Inteligencia Artificial. Otra interrogante la encontramos en la reparación del daño y en la pregunta: ¿Cómo calcular la reparación del daño por una afectación que surge de un sistema autónomo que no fue programado ni supervisado por un ser humano? ¿El ser humano que está más cercano a la tecnología debe de reparar el daño o es la persona jurídica la que debe hacerse responsable? ¿Es un caso que puede solucionarse basado en el modelo tradicional de daños y perjuicios? ¿tendría que ser evaluado como una afectación a los Derechos Humanos de las personas?

La **problemática** puede analizarse desde diferentes enfoques, pero para los fines de esta investigación hemos decidido acotarla a la siguiente reflexión: los sistemas autónomos, creados para auxiliar a seres humanos en la toma de decisiones (y que actualmente muchos ya tienen la capacidad de funcionar sin la participación del ser humano), pueden causar daños o afectar Derechos Humanos, su regulación no está contemplada con precisión en los sistemas jurídicos, ni tampoco existen órganos especializados que realicen su supervisión y control. Esto se debe revisar de inmediato ya que los sistemas autónomos están presentes en nuestra vida cotidiana y están guiando nuestras decisiones. La comunidad jurídica

no está vinculada a estos temas con la profundidad que amerita, lo ve en un nivel de análisis que no permite la construcción de propuestas para un cambio en la enseñanza y aplicación del derecho que contribuya a la construcción de un ecosistema de Inteligencia Artificial (IA). Adicional a ello, existe una escasez de operadores jurídicos que tengan las habilidades para entender, analizar, tomar decisiones y emitir resoluciones (peritos, jueces, fiscales, árbitros, profesores universitarios, entre otros) con el conocimiento técnico especializado para garantizar una reparación del daño a las víctimas, la no repetición del acto y la protección de los Derechos Humanos.

Para el desarrollo de la problemática, a la par nos hacemos varios cuestionamientos: ¿En qué momento la IA llegó a este punto conflictivo? ¿De qué tecnología estamos hablando? ¿Toda la IA afecta Derechos Humanos? ¿Qué entienden los abogados por IA? ¿Los operadores jurídicos (fiscales, jueces, peritos, abogados, etcétera) saben diferenciar entre la IA controlada por seres humanos y la IA no controlada por seres humanos? ¿A quién le corresponde el pago de la reparación de los daños ocurridos sino se puede ubicar el acto volitivo que genera la responsabilidad? ¿Las categorías de responsabilidad contenidas en la legislación civil sirven para la reparación de daños y perjuicios? ¿Las categorías de responsabilidad contenidas en la legislación penal permiten reparar el daño a las víctimas?

La problemática se complica cuando la IA se involucra con actividades humanas específicas, como, por ejemplo, reclutamiento de personal, otorgamiento de créditos o préstamos, ingreso o permanencia en sistemas escolares, entrega de beneficios sociales; emisión de resoluciones judiciales, diagnóstico y aplicación de tratamientos médicos, etcétera. El caso de estudio que hemos elegido para esta investigación se ubica en el campo de los servicios de salud, es un sistema de IA (*Watson Health*), que toma decisiones para auxiliar a un ser humano (el médico) en el diagnóstico y/o tratamiento de alguna enfermedad (cáncer) y que su uso puede causar responsabilidad médica y, hasta una posible violación a Derechos Humanos.

Quizás estemos entrando en una etapa de construcción de una nueva categoría de responsabilidad jurídica (o moral) derivada del uso de la IA y de la

ausencia de: a) la presencia física (de personas y/o de evidencias), b) la carencia de actos volitivos generados por las personas; o c) del entendimiento del comportamiento humano basado en hechos generados o causados por otro ser humano.

Las propias definiciones de “inteligencia”, “conciencia”, “cerebro”, “mente” y sus funciones no han podido ser estudiadas con la profundidad debida para llegar a un uso clarificador tanto para la IA como para la Ingeniería, Filosofía y el Derecho (tan sólo el concepto de libre albedrío y de agencia moral siguen siendo sometidos a amplios debates).

En cuanto a los fines del derecho y de las diferentes teorías de la responsabilidad que existen en el mundo occidental, creemos que sus principios y uso pueden quedar limitados o superados para muchos fenómenos sociales que están surgiendo por el uso de la IA y que el trabajo de investigación académica de ingenieros, filósofos, científicos de datos, médicos, psicólogos, entre otros, deben construirse de forma transdisciplinaria.

Si los sistemas de IA van a tomar decisiones a semejanza del ser humano, entonces, éstas no deben estar alejadas de la Moral (o de la Ética) y de los principios del Derecho. La IA debe ser parte de un ecosistema de protección social con un soporte jurídico y con componentes morales (éticos) tanto para emitir sus decisiones como para la revisión de estas, ya sea por un ser humano o por otro sistema de IA que ayude al ser humano a evaluar y proteger sus derechos.

A lo largo de la investigación se demuestra que no todos los sistemas de IA que ayudan a la toma de decisiones son un riesgo para el ser humano, sólo aquellos cuyos algoritmos son considerados “cajas negras” y cuyas entradas (datos) y las salidas (soluciones) las conocemos pero no las operaciones algorítmicas que ocurren para llegar a los resultados, no su trazabilidad, que le permita al ser humano entender la responsabilidad que genera su creación o uso de este tipo de IA, diluyéndose los elementos tradicionales de la responsabilidad civil o penal y complicándose la aplicación de alguna sanción o punición. Este tipo de sistemas derivan del uso de Machine Learning – Deep Learning y del cómputo cognitivo y son utilizados en diversas industrias que aún los siguen considerando como un

instrumento, dispositivo o software tradicional. Esto debe cambiar en beneficio de ese ecosistema de protección social con soporte jurídico y ético (moral).

Evidentemente, mucha de la responsabilidad está en la empresa y en su gobierno corporativo. Para efectos de esta investigación, entendemos a la empresa como una comunidad de personas que se unen para ciertos fines y que su convivencia cotidiana “acarrea responsabilidades”. Esta definición llanista, conlleva también a entender que la unión de voluntades y la obtención de beneficios, debe ser medida con base en los riesgos legales, tecnológicos y éticos y la responsabilidad de las personas.

La investigación analiza las normas que regulan la responsabilidad jurídica y su relación con la responsabilidad moral. En esta segunda dimensión también aparecen cuestionamientos hacia los principios morales y el uso de la ética en las empresas que utilizan las nuevas tecnologías. Por ejemplo, ¿qué principios morales (éticos) deben incorporarse a los sistemas de IA? y ¿qué tipo de Moral (Ética) debe incorporarse a la empresa? La respuesta a la primera pregunta está en los Derechos Humanos que usan como parámetro de medición del comportamiento de un sistema inteligente en el Derecho y debe ser sumada a las aportaciones de filósofos o eticistas que están explorando nuevas definiciones como la de “Agente Moral Artificial Autónomo” y la “Ética del comportamiento de los sistemas de IA” (o Ética por diseño).

La segunda pregunta, nos lleva a los postulados de responsabilidad moral y la integridad corporativa. Aquí, se utilizan los postulados de responsabilidad moral de Carlos Llano Cifuentes, con la advertencia de que no se busca entrar al desarrollo profundo de la obra llanista, ni mucho menos a la búsqueda de conceptos filosóficos para una dialéctica argumentativa de dicho autor, ya que esto sobre pasa a los fines de esta investigación. Pero sí se hace una revisión de los principales textos en los que Carlos Llano desarrolla el concepto de la responsabilidad moral, su relación con la responsabilidad jurídica y la irrupción del internet como factor que ha venido modificando las relaciones humanas, pues se considera que es un autor moderno, que tuvo la fortuna de estudiar filosofía, ejercer la docencia, y a la par, hacerse cargo de diversas funciones directivas en el sector empresarial. Este perfil

no es común en los tiempos que vivimos, ni en México ni en otras latitudes. Por eso también la recomendación a su estudio y uso de sus aportaciones.

Entre las aportaciones de este trabajo de investigación está la propuesta del modelo de metodología para la medición de riesgos por impactos adversos a las empresas, basado en tres dimensiones: legal, tecnológica y ética. En su dimensión ética, se utiliza el pensamiento de Carlos Llano, pero puede utilizarse cualquier otro que haya estudiado el tema de la responsabilidad moral y cuya concepción dirija el actuar empresarial basado en la persona, su dignidad y la razón directiva.

¿Cómo garantizar a la sociedad que los sistemas de IA no tomen decisiones prohibidas por la ley y que a su vez sean respetuosos de los principios morales o éticos? Si los seres humanos no pueden determinar por sí mismos que las decisiones o los actos fruto de estas decisiones son respetuosos de Derechos Humanos y no pueden saber si son resultado de la programación humana o de algún humano, entonces quizás, se requiere la intervención de otro sistema inteligente construido en un segundo nivel que garanticen la no violación de las leyes ni de los principios morales contenidos en cada uno de los Derechos Humanos.

Una vez ubicado el responsable: ¿cuál sería el camino para categorizar la responsabilidad del propietario, usuario, diseñador, fabricante o máquina que utiliza a la IA para decidir ella misma? ¿La ley vigente regula los supuestos en los que se generó un daño por el uso de sistemas de IA? ¿Qué instituciones, normas, organizaciones privadas están funcionando o requieren funcionar como parte de un ecosistema de protección social para el uso de IA?

Esta investigación busca contribuir a la reflexión de estos temas novedosos y propone algunos cimientos para la construcción de un “Ecosistema de protección social para el uso de IA” (o quizás varios ecosistemas regionales). Dicho ecosistema requiere de: a) normas morales y jurídicas (dentro y fuera de la empresa); b) agencias reguladoras expertas en IA, unas de corte público, responsables de capacitar, crear políticas, procedimientos y procesos para dirimir controversias, llevar un registro público de algoritmos; supervisar el impacto social y ambiental de esta tecnología; también, agencias que tengan facultades de supervisión y sanción

en aquellos casos que se compruebe una afectación a Derechos Humanos; c) empresas certificadoras y auditoras de algoritmos para que midan los posibles riesgos legales de las empresas que utilizan IA y se sumen a los riesgos de la organización; d) Programas y responsables de implementar el *Compliance* al interior de las empresas que crean o utilizan la IA; e) Creación de metodologías para la medición y evaluación del impacto adversos de la IA en los Derechos Humanos; f) Supervisiones periódicas por las agencias reguladoras del gobierno; g) Tribunales especializados en dirimir controversias por temas de responsabilidad algorítmica y afectación a Derechos Humanos por el uso de este tipo de tecnología; h) Creación de un comité de IA en la estructura de la Organización de las Naciones Unidas (ONU) y sus homólogas regionales, como funciones para casos confirmados de violación de Derechos Humanos por uso de IA, con estructura similar a la de otros comités que existen actualmente; i) actualización de planes de estudio en escuelas, universidades y centros de investigación para incluir los nuevos conceptos y usos de la IA y hacer labores de prevención y de divulgación; j) Participación de los organismos empresariales y cámaras de comercio en la capacitación y divulgación de estos temas; k) utilizar los beneficios de la Ética empresarial y de la participación de filósofos expertos en uso de IA; l) el desarrollo de políticas, procesos, procedimientos, sistemas, certificaciones y mecanismos de auditoría o vigilancia al interior y exterior de las empresas que generan IA o de los gobiernos que la estén utilizando para alguna actividad en la que se vea involucrada la población.

La problemática es poliédrica y supera los fines de una investigación jurídica, por lo que se debe de estudiar este fenómeno de forma transdisciplinaria, con los contenidos de la norma jurídica vigente, los estándares de diversas organizaciones que han estudiado el tema desde un enfoque tecnológico y la Ética empresarial.

Como **objetivo general**, la investigación busca analizar los diferentes sistemas de Inteligencia Artificial que han venido utilizándose en el mundo, ubicando aquellos que fueron creados para auxiliar al ser humano en la toma de decisiones y su uso en el campo de la medicina (para el diagnóstico y/o tratamiento de enfermedades como el cáncer) con el fin de conocer sus orígenes, evolución, principios básicos de funcionamiento y las consecuencias o problemas que están

generando actualmente para vincular su uso con las categorías de la responsabilidad jurídica y el uso de la Ética empresarial. Todo ello se utilizará para demostrar como actualmente no es fácil para los jueces, fiscales, abogados, peritos o cualquier experto en temas de responsabilidad legal, determinar si una persona es o no responsable por el uso de un sistema autónomo creado para auxiliar al ser humano en la toma de decisiones. La responsabilidad se iría hacia el ser humano que se encuentra más cercano a la tecnología que generó la decisión, sin participación del mismo en ella.

La IA y su uso por los gigantes de la tecnología como Google, Amazon, Apple, IBM y Microsoft aparecen en la vida cotidiana de las personas, generando un nuevo orden mundial basado en “los datos” y en el control que tienen estas empresas (que supera las posibilidades de cualquier Estado) sobre las decisiones, acciones, pensamientos y salud de las personas. Un ejemplo es la llamada *Healthcare technology* o *Health technology* definida por la Organización Mundial de la Salud (OMS) como “la aplicación de conocimientos y habilidades organizados en forma de dispositivos, procedimientos, sistemas, medicamentos y vacunas desarrollados para salvar un problema de salud y mejorar la calidad de vida”¹ y en

¹ En 2007, la LX Asamblea Mundial de la Salud, se reconoció que las tecnologías sanitarias ofrecen a los prestadores de servicios de salud herramientas necesarias para la prevención, el diagnóstico, el tratamiento y la rehabilitación efectivos y eficientes. Entendiendo que las tecnologías sanitarias, en particular los dispositivos médicos, representan un desafío económico y técnico para los sistemas de salud de muchos estados, y preocupado por el desperdicio de recursos resultante de inversiones inapropiadas en tecnologías sanitarias, en particular dispositivos médicos, que no satisfacen necesidades de alta prioridad, son incompatibles con las infraestructuras existentes, se utilizan de forma irracional o incorrecta, o no funcionan de manera eficiente; Existe una necesidad de que los estados y los donantes frenen los crecientes costos estableciendo prioridades en la selección y adquisición de tecnologías sanitarias, en particular de dispositivos médicos, sobre la base de su impacto en la carga de morbilidad, y de asegurar el uso eficaz de los recursos a través de planificación, evaluación, adquisición y gestión; Se requiere ampliar los conocimientos especializados en el campo de las tecnologías sanitarias, en particular los dispositivos médicos; La OMS pide a sus miembros: 1) recopilar, verificar, actualizar e intercambiar información sobre tecnologías sanitarias, en particular dispositivos médicos, como ayuda para la priorización de necesidades y la asignación de recursos; 2) formular, según proceda, estrategias y planes nacionales para el establecimiento de sistemas de evaluación, planificación, adquisición y gestión de tecnologías sanitarias, en particular dispositivos médicos, en colaboración con el personal involucrado en la evaluación de tecnologías sanitarias y la Ingeniería biomédica; 3) elaborar directrices nacionales o regionales de buenas prácticas de fabricación y reglamentarias, establecer sistemas de vigilancia y otras medidas para garantizar la calidad, seguridad y eficacia de los productos sanitarios y, en su caso, participar en la armonización internacional; 4) establecer, cuando sea necesario, instituciones regionales y nacionales de tecnología sanitaria, y colaborar y crear

donde se están invirtiendo miles de millones de dólares para la creación de sistemas autónomos para la toma de decisiones y dispositivos para mejorar la salud de las personas.

Se acotó el tema de la investigación exclusivamente al uso de los sistemas de IA que se utilizan para la toma de decisiones médicas (Sistemas Expertos Decisionales y/o Machine Learning – Deep Learning), eligiendo a IBM -uno de los nueve gigantes de la tecnología, el más longevo y robusto- y su sistema *Watson* ya que es considerado como el pionero de esta tecnología y el mejor evaluado en el cuadrante Gartner.

IBM, es pionero en el uso de tecnología para la toma de decisiones basada en computación cognitiva. Por dar solo algunos datos: *IBM Services* diseña, construye y ejecuta los sistemas y servicios fundamentales que son la columna vertebral de la economía mundial. Los clientes a los que IBM da soporte incluyen: cuatro de las cinco principales aerolíneas por ingresos; ocho de cada 10 operadores móviles líderes, ocho de cada 10 mayores fabricantes de automóviles, cuatro de los 10 minoristas más grandes del mundo; y, siete de cada 10 compañías de seguros más grandes.²

Se parte de la siguiente **hipótesis**: El uso de algunos modelos de IA que auxilian al ser humano a la toma de decisiones (en este caso, decisiones médicas), están produciendo afectaciones a las personas, algunas de estas afectaciones pueden causar daños cuya gravedad está regulada de forma poco clara en la legislación; dichas afectaciones, a la vez pueden generar daños y/o impactos adversos a Derechos Humanos, trasladando la responsabilidad al humano que se encuentre más cercano a la tecnología. Sin embargo, ni la doctrina ni la legislación

asociaciones con los proveedores de atención de la salud, la industria, las asociaciones de pacientes y las organizaciones profesionales, científicas y técnicas; 5) recopilar información que interrelacione los dispositivos médicos, que se ocupen de las condiciones prioritarias de salud pública en diferentes niveles de atención y en diversos entornos, con la infraestructura, los procedimientos y las herramientas de referencia necesarios. Cfr. World Health Organization, WHA60.29 *Health technologies, Sixtieth World Health Assembly*, Ginebra, 14 a 23 mayo de 2007, Resolutions and Decisions Annexes. Consúltase en: https://www.who.int/healthsystems/WHA60_29.pdf?ua=1

² BNP Paribas Signs an Agreement with IBM Services to further deploy its Cloud Strategy, Armonk, N. York, 22 de enero de 2019. Consúltase en: <https://newsroom.ibm.com/2019-01-22-BNP-Paribas-Signs-an-Agreement-with-IBM-Services-to-further-deploy-its-Cloud-Strategy>

vigente en México, contempla alguna solución con la claridad necesaria para categorizar la responsabilidad de las personas, de la empresa (o de las máquinas). Tampoco queda claro cómo utilizar las aportaciones de la Ética empresarial para prevenir riesgos legales y tecnológicos.

En este sentido, las empresas de IA, como cualquier empresa, deben ser entendidas como una comunidad de personas que, a través de sus directivos, toma decisiones. Estas decisiones a veces pueden ser certeras y exitosas y a veces pueden ser erróneas y dañinas (generando impactos adversos a la sociedad, al medio ambiente a su gobierno corporativo o a sus *stakeholders*). Cuando una o varias empresas, a través de sus directivos, deciden utilizar o poner al servicio de sus clientes un sistema de IA, deben establecer mitigantes de tres tipos:

- a) Mitigantes jurídicos, que deben ser tomados del sistema constitucional o convencional, para prevenir riesgos relacionados con la legislación del sector en el que participa la empresa.
- b) Mitigantes tecnológicos, basados en el tipo de Inteligencia Artificial que se utiliza y tomando en cuenta las directrices o guías que han emitido las asociaciones de ingenieros expertas en estos temas y sumarlas a los controles jurídicos, para prevenir riesgos por el uso de las nuevas tecnologías.
- c) Mitigantes éticos, para mitigar riesgos derivados del comportamiento humano y del uso de la tecnología que permitan crear una cultura basada en la integridad y en el fomento de valores, principios y reglas de comportamiento de las personas, organizaciones y, en algún momento, de los sistemas autónomos.

En todo esto, pensemos dos temas de relevancia social. El primero: Un sistema autónomo para la toma de decisiones que por alguna razón se equivoca, sin la participación del ser humano en esa equivocación, responsabiliza al ser humano mismo ya que el sistema jurídico y sus operadores están entendiendo que la tecnología es solo un instrumento, dispositivo o software, y como tal, el ser humano que se encuentra más cercano a la tecnología es el responsable. No hay

un juicio de reproche soportado en norma jurídica que guíe la responsabilidad hacia la agencia artificial que produjo el error.

El segundo, ¿Qué decisión tomaría usted al saber que su vida, su salud o su destino gira en torno a las decisiones que toma un sistema de IA sin intervención humana? y ¿Qué haría si con motivo de sus actividades profesionales o en algún viaje vacacional se produce un daño causado por un sistema de IA del cual se le responsabiliza? Y, adicionalmente, a sabiendas de que usted no hizo ningún acto volitivo ni tuvo que ver con el daño, le piden que demuestre su inocencia a través de la “trazabilidad algorítmica”, ya que la ley no contempla ningún tipo de responsabilidad para agentes autónomos que toman decisiones sin intervención humana por considerarlos un dispositivo o *software*. Piense también que usted, sus familiares, amigos y compañeros de trabajo utilizan esta tecnología de manera cotidiana y en todos lados (en su casa, oficina, auto, comercios, restaurantes, hoteles, aeropuertos, trámites gubernamentales y bancarios) y que el daño causado se puede replicar tal como hoy ocurre un accidente de tránsito o aéreo.

Lamentablemente, en todo esto, Los operadores de los sistemas de justicia (penal, civil, administrativo, laboral, entre otros) no están dimensionando la problemática, tampoco los legisladores, ni abogados. Por tanto, los sistemas de justicia mantienen sus estándares de argumentación y razonamiento probatorios sin la posibilidad de reducir la responsabilidad del ser humano más cercano a la tecnología (y transferirla a los sistemas autónomos o contemplarlos como una eximente de responsabilidad). El costo social y de afectación a las personas será injusto e inhumano ya que la responsabilidad de los daños o afectación a derechos humanos recaerá en la persona que no tuvo posibilidad de tomar una decisión para controlar al sistema autónomo sin tomar en cuenta que el sistema le restó capacidades de análisis, reacción y solución de la problemática.

Para verificar la anterior afirmación y como parte de los **objetivos particulares**, investigación hace un recorrido por las soluciones que ofrece la legislación civil y penal, los criterios jurisprudenciales de la Suprema Corte de Justicia de la Nación y los Tribunales Colegiados; así como a las aportaciones filosóficas de Virginia Dignum y su texto *Responsible Artificial Intelligence. How to*

Develop and Use AI in a Responsible Way; sus reflexiones complementan las aportaciones de Carlos Llano Cifuentes.

Acudimos también a pensadores como Hans Jonas, H. Hart, Donald Dworkin, y dos mexicanos excepcionales: Ernesto Gutierrez y González y Carlos Llano Cifuentes. Todos, por su cuenta, nos llevan a entender las diversas discusiones que han existido en torno al concepto responsabilidad, responsabilidad moral, responsabilidad jurídica, responsabilidades de la empresa, responsabilidad médica y responsabilidad algorítmica.

La **delimitación temporal** del tema fue pensada con base en los orígenes del concepto Inteligencia Artificial (1956-1957). La revisión hecha se enfoca a determinar los distintos usos del término, tipos, clasificaciones, modelos algorítmicos, precursores y principales novedades que existen actualmente. En esa revisión encontramos múltiples usos de la IA, por lo que se decidió hacer una **delimitación material del tema** a un área de la IA específica que tiene que ver con el uso de sistemas inteligentes auxiliares para la toma de decisiones humanas y su desarrollo en el campo de la medicina. A la vez hacemos una revisión de las normas y orígenes de la responsabilidad moral, jurídica y los cuatro tipos responsabilidades de Carlos Llano Cifuentes. Como una conclusión extra, creemos que el Instituto de Humanidades y el Instituto Panamericano de Alta Dirección, de la Universidad Panamericana, deben abrir líneas de investigación y participar en el desarrollo de estos temas, articulando cuerpos de investigación y docencia para formar directivos conocedores de las responsabilidades que existen por el uso de la IA la integridad en la empresa.

En muchas partes de América Latina, los gobiernos, empresas y universidades están evolucionando a pasos más lentos de lo esperado, lo cual es preocupante porque se requieren de muchos especialistas que estén aportando ideas para el análisis y la actualización de los sistemas jurídicos, sobre todo en el uso de estas tecnologías y los nuevos conceptos que se deben utilizar para entender la responsabilidad de la empresa. Lo cual no significa que nunca nadie

haya visto el problema del Derecho y la tecnología³; simplemente queremos manifestar que de 2019 a 2022 (periodo de la redacción e incorporación de ajustes de la investigación) no encontramos algún experto con formación jurídica que estuviera elaborando una investigación transdisciplinaria (Ingeniería - Filosofía - Derecho).

También hay que manifestar que es poco común que los estudiosos de la informática intenten vincular sus conocimientos con las aportaciones de la Teoría del Derecho o de la Teoría de la Responsabilidad Jurídica. Ingeniería y Derecho se habían mantenido desvinculados en sus construcciones teóricas y metodológicas, sin hacer una transversalidad de conocimientos; cada uno haciendo aportaciones con su propio método y su propia especialización. Salvo algunas excepciones⁴, las universidades mantienen una dinámica académica muy enfocada al Derecho creado en los albores del siglo XIX. Sin embargo, el desarrollo vertiginoso de la tecnología hace insuficiente estos esfuerzos académicos y se requiere un perfil de profesionistas con formación transdisciplinaria para el análisis y la atención de los

³ Quizás el primer autor que intentó poner las bases en el tema de la Ética y uso de tecnología fue Vittorio Frosini, en su libro *Cibernetica, diritto e società*, Edizione di Comunità (Milano, 1968) haciendo un conjunto de reflexiones sobre el impacto tecnológico, los problemas éticos y jurídicos tradicionales. Frosini es considerado el padre del Derecho Artificial en Italia y la creación del concepto “libertad informática” como un nuevo derecho, con el significado de autotutela de la propia identidad informática, es decir, el derecho de controlar (accesar, rectificar o cancelar) los datos personales inscritos en un banco de datos. Existe una edición en español actualizada de Ediciones Olejnik, Santiago, 2019 (ver capítulo V Ética y Ciberética). La cibernética, etimológicamente es la “ciencia de la decisión”. En México, Julio Tellez Valdés, ha impulsado el Derecho Informático desde 1983 con su texto *La Informática y el Derecho* publicado por el INEGI y su actual obra *Derecho informático*, Quinta Edición, México, McGraw Hill, 2021.

⁴ La Universidad Panamericana está construyendo un proyecto muy interesante en el que participan alumnos y maestros de las facultades de Filosofía, Ingeniería y Derecho. Agradecemos a Karen González Fernández y a Enrique Siqueiros la posibilidad de incorporarnos a esas actividades. Juan G. Corvalán, director del Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires, ha conseguido subir a la Filosofía a la Ingeniería y al Derecho con financiamiento internacional para formar especialistas en IA (<https://www.uba.ar/noticia/20195>) La Universidad Autónoma de Chile ha decidido crear un programa para estudiantes en el que unen los conocimientos de las carreras de Derecho e Ingeniería (2019). El programa de “Minor en IA y Derecho” busca desarrollar conocimientos y habilidades “conectadas con la programación, el uso de algoritmos, el diseño de servicios o productos digitales para la resolución de problemas jurídicos y con dominio de sus límites normativos, con relación al plano ético, al de protección a la privacidad y los datos personales y al de seguridad de activos de la información”. Entre las materias que se imparten destacan: Ética e Inteligencia Artificial; Diseño de Aplicaciones; Protección de datos personales; Ciberseguridad; Aplicaciones de IA. Cfr. <https://uautonoma.cl/wp-content/uploads/2018/10/Di%CC%81ptico-Minor-IAAd.pdf> [20 de julio de 2022]

problemas que surgen por la creación y utilización de algoritmos de *Machine Learning - Deep Learning*.

Debemos estar conscientes que la evolución de la IA se ha logrado gracias a la fusión de muchos campos de estudio que inician con la Informática y la Ingeniería Electrónica pasando por las Matemáticas, la Lógica, la Estadística, la Psicología, la Biología, la Lingüística, la Filosofía y el Derecho (con el cual ahora se están construyendo varias normas para la protección de derechos).⁵ Si lo entendemos de esta manera, un problema que se genera por el uso del ML o DP debe ser investigado y solucionado con varios enfoques y varios métodos; los métodos a su vez pueden ayudar a encontrar varios problemas y sus posibles soluciones, que no se podrían vislumbrar sin una transdisciplinariedad.

Intentar construir estudios transdisciplinarios de la IA es una realidad que comienza a tomar una dinámica novedosa. Si bien es cierto que nuestro país ya “...tiene una oferta muy amplia de grados, postgrados, sedes, programas, laboratorios y centros de investigación que ofrecen formación en temas relacionadas con la IA”⁶; que “...la investigación sobre estos temas es prolífica y existen iniciativas colaborativas con el sector empresarial, y también con los gobiernos federal y local”⁷; y que “se están explorando soluciones innovadoras a retos sociales de manera colaborativa con aquellas organizaciones de la sociedad civil que buscan aprovechar las nuevas herramientas para ampliar su impacto”⁸;

⁵ La informática y la Ingeniería eléctrica determinan cómo se implementa la IA en el software y el hardware. Las matemáticas y las estadísticas determinan modelos viables y miden el desempeño. Debido a que la IA se basa en cómo creemos que funciona el cerebro, la psicología y la lingüística desempeñan un papel esencial en la comprensión de cómo podría funcionar la IA. Y la Filosofía proporciona orientación sobre la inteligencia del ser humano y diversas consideraciones Éticas. Si bien la versión de ciencia ficción de la IA puede ser una posibilidad distante, ya vemos más y más IA involucrada en las decisiones que tomamos todos los días. Desde 1956, la IA ha demostrado ser útil en diferentes campos del saber, impactando la vida de las personas y nuestra sociedad de manera significativa. ML es la herramienta que más impacto ha tenido.

⁶ Un estudio sobre el estado del arte de la IA en América Latina lo encontramos en: Constanza Gómez Mont; Claudia May Del Pozo; Cristina Martínez Pinto; Ana Victoria Martín del Campo Alcocer (C Minds), *La Inteligencia Artificial al Servicio del Bien Social en América Latina y el Caribe: panorámica regional e instantáneas de doce países*, México, Banco Interamericano de Desarrollo, 2020. Cfr. https://7be0258a-f55e-4968-87a4-4c067a3d1160.usrfiles.com/ugd/7be025_bbfdb0a0224c43a2ad371b87b66c2eae.pdf [20 de julio de 2022]

⁷ Ibidem., p. 96.

⁸ Idem.

también lo es, que el Derecho, como conjunto de normas y principios que regulan la vida del hombre en sociedad y sus problemas, aún no está presente con la amplitud y profundidad necesarias para generar soluciones.

Un dato curioso es que México es ubicado a nivel internacional como uno de los primeros países que se ha dado a la tarea de crear una “Agenda Nacional de Inteligencia Artificial”⁹ y existen muchas personas que están involucrándose en estos temas. Sin embargo, no se ha hecho nada para su materialización y la producción de estudios, artículos, ensayos o conferencias aún no se dirige hacia la construcción de un ecosistema de cuidado social para el uso de IA o por lo menos para la creación de un modelo de certificación y elaboración de auditorías algorítmicas con un enfoque basado en los Derechos Humanos, como sí ocurre en varios países de Europa.

Durante cuatro años de investigación, se dio seguimiento a conferencias, seminarios o congresos en los que se discutieron temas sobre IA, Ética, Filosofía y Derecho, para conocer las posturas de la comunidad académica y los intentos de regulación existentes. En México existen varios espacios de interlocución que tratan estos temas y se organizan anualmente para tratar a la IA¹⁰, pero en ninguno

⁹ En abril de 2018 se hizo el documento “Towards an AI Strategy in Mexico: Harnessing the AI Revolution”, patrocinado por la embajada del Reino Unido en México y elaborado por CMINDS (Cfr. <https://www.oxfordinsights.com/mexico>). Posterior se hizo la “Agenda Nacional de Inteligencia Artificial para México” que se ha venido actualizando y fortaleciendo por CMINDS. Puede consultarse en: <https://www.ia2030.mx/> La presentación de la Agenda Nacional de IA2030 de México se organizó por C Minds en el marco de la Coalición IA2030Mx y patrocinado por NVIDIA (2 de septiembre 2020). C Minds es una organización muy activa, a nivel nacional e internacional, con especialistas que están analizando diversos temas y con varios documentos de consulta en: <https://www.cminds.co/reports> [20 de julio de 2022]

¹⁰ En Latinoamérica se encontraron discusiones que no entran al fondo del uso del ML y se mezclan con la robótica y sobre todo con historias producidas por la ciencia ficción, en donde los seres humanos y su toma de decisiones son superados por las máquinas. La UNAM, a través del Instituto de Investigaciones Jurídicas, intentó vincular a la comunidad jurídica con estos temas y el trabajo de sus investigadores. En mayo de 2020, se organizó la primera sesión del Seminario “Inteligencia Artificial en Respuesta al COVID-19”. Consúltase el temario en: <https://www.juridicas.unam.mx/actividades-academicas/2401-seminario-lidia-primera-sesion-2020> El Colegio Nacional (México) organizó un ciclo de conferencias para reflexionar sobre “La humanidad y la máquina”. No se encontró una visión precisa sobre lo que es el uso de la IA (en cuanto al ML), el Derecho y la regulación del uso de algoritmos. Destaca la sesión del 28 de agosto de 2020 en la que se trataron temas por especialistas con formación jurídica (pero sin una especialización sobre ML). El coloquio puede consultarse en: <https://www.youtube.com/user/elcolegionacionalmx/videos> La Sociedad Mexicana de Inteligencia Artificial organizó el panel “Implicaciones Éticas y filosóficas del uso de los avances recientes en inteligencia artificial y el aprendizaje máquina”. En IIMAS-UNAM

encontramos especialistas que estudiaran la problemática con el enfoque conceptual que se encuentra en esta investigación.

Algunos ejemplos de eventos relevantes los encontramos en el *IEEE* (*Institute of Electrical and Electronics Engineers*)¹¹ y en la ACM (*Association of Computing Machinery*)¹². En el evento más cercano a estos temas se hizo en Barcelona (enero, 2020), la ACM puso como eje de la discusión el estudio de los sistemas algorítmicos que se están adoptando en un número creciente de ámbitos del quehacer humano, la responsabilidad que genera y los mecanismos de auditabilidad algorítmica. Los sistemas de inteligencia artificial que se utilizan filtran, clasifican, registran, recomiendan, personalizan y simulan la experiencia humana, y cada vez toman más decisiones con mayor impacto en el acceso a la educación, oferta de empleos, créditos, seguros, servicios de salud, justicia penal (libertad condicional), seguridad social y asuntos migratorios.

Entre las reflexiones que se hicieron fue que dichos sistemas aportan innumerables beneficios, pero a su vez, entrañan riesgos inherentes, como la codificación y el afianzamiento de prejuicios que pueden afectar la dignidad de la persona o sus derechos, la reducción de la rendición de cuentas y la obstaculización de las garantías procesales; también aumentan la asimetría de la información entre las personas cuyos datos alimentan estos sistemas y los grandes actores capaces de inferir información potencialmente relevante.¹³

se realizó el seminario “El rol de la IA en la sociedad”. La Escuela Libre de Derecho tuvo un “Seminario de Inteligencia Artificial”. Estos eventos fueron registrados en la *Agenda Nacional Mexicana de Inteligencia Artificial*, México: CMinds. <https://www.ia2030.mx/> La Academia Nacional de Ciencias de Buenos Aires, a través del Centro de Estudios Tecnológicos Inteligentes organizó un ciclo de conferencias sobre “Sistemas Inteligentes y Ética” que se pueden consultar en: <https://www.youtube.com/channel/UC3UzBZL-5VRkqoFYinf4kGg> [20 de julio de 2022]

¹¹ En la primavera de 1884, un pequeño grupo de personas de profesiones eléctricas se reunió en Nueva York, EE. UU. Formaron el Instituto Americano de Ingenieros Eléctricos. En octubre, la AIEE celebró su primera reunión técnica en Filadelfia, EE. UU. El 1 de enero de 1963, la AIEE se fusiona con el Instituto de Ingenieros de Radio (*Institute of Radio Engineers, IRE*) para formar el Instituto de Ingenieros Eléctricos y Electrónicos, o IEEE. En 2010, IEEE estaba compuesto por más de 395,000 miembros en 160 países. <https://www.ieee.org/about/ieee-history.html#foundation-of-the-aiee> [20 de julio de 2022]

¹² Todos los eventos de la *Association of Computing Machinery* ACM FAccT se encuentran en: <https://facctconference.org/2020/programschedule.html> [20 de julio de 2022]

¹³ Cfr. *Association for Computing Machinery, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York, 2020. Consúltese en: <https://dl.acm.org/doi/proceedings/10.1145/3351095> [20 de julio de 2022]

Resulta importante decir que, durante 2018 se presentaron varios casos polémicos derivados del uso de la IA, en específico, encontramos varios ejemplos que ocurrieron por el uso de inteligencia artificial:

1. La violación de datos privados de Facebook sobre más de 50 millones de usuarios a una empresa de marketing político contratada para la campaña presidencial de Donald Trump, revelada en marzo de 2018;
2. El contrato de Google con el Pentágono para que el software de visión computarizada se utilice en zonas de combate, revelado en marzo de 2018;
3. La venta de tecnología de reconocimiento facial por parte de Amazon a los departamentos de policía, revelada en mayo de 2018;
4. El contrato de Microsoft con la Inmigración y Control de Aduanas de los Estados Unidos, revelado en junio de 2018; y
5. La colaboración secreta de IBM con el Departamento de Policía de Nueva York para el reconocimiento facial, revelada en septiembre de 2018.¹⁴

Tanto usuarios como trabajadores, organizaciones de la sociedad civil y gobiernos están bajo la mira y se han organizado compañías tales como #TechWontBuildIt, #NoTechForICE, #Data4BlackLives, para exigir que el uso una IA sin discriminación, sin violación a la privacidad o sin generar una afectación a Derechos Humanos. Esto ha traído consecuencias, hoy por ejemplo en muchas partes del mundo el reconocimiento facial ya fue suspendido y su uso es para casos relevantes y necesarios relacionados con temas de seguridad nacional o protección de instituciones públicas.¹⁵

Esta problemática nos lleva a investigar el tema y cómo está interviniendo los estados a través de sus sistemas de justicia, existiendo varias preguntas: ¿Cómo los abogados, jueces, mediadores, árbitros, usuarios pueden exigir la reparación del daño por un sistema de IA si ni siquiera saben lo que es o cómo funcionan sus algoritmos? ¿Cómo buscar una regulación o, quizás hasta una redefinición de conceptos como equidad, transparencia, responsabilidad, privacidad, dignidad si se

¹⁴ Cfr. Rodrigo, Ochigame, "The Invention of Ethical AI How Big Tech Manipulates Academia to Avoid Regulation" en *The Intercept*, 20 de diciembre, 2019. <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence> [20 de julio de 2022]

¹⁵ *Idem*.

desconoce lo que pueden hacer y lo que no pueden hacer? ¿De qué manera se auditaría a una empresa que utiliza esta tecnología para verificar si es responsable o no por el comportamiento descontrolado de sus algoritmos y los sesgos que genera al tomar decisiones? ¿Existe alguna autoridad reguladora o supervisora que pueda intervenir en estos temas para defender los derechos de las personas afectadas? ¿Cómo categorizar la responsabilidad civil, penal o administrativa por el uso de sistemas inteligentes? ¿Qué conceptos cambian o se deben añadir a la Teoría de la Responsabilidad Jurídica? ¿De qué manera intervienen otras disciplinas para ayudar al Derecho a resolver una violación de Derechos Humanos por el uso de IA? ¿Qué pasa con conceptos como el de responsabilidad, autonomía de la voluntad, libre albedrío, agencia moral si el que toma la decisión no es precisamente un ser humano? ¿Cómo reducir los sesgos algorítmicos que afectan la dignidad, la privacidad, la no discriminación, la igualdad de género, la implementación de políticas públicas? ¿existe la posibilidad de construir un ecosistema de protección social? ¿el modelo de responsabilidades de Carlos Llano Cifuentes se puede utilizar para las empresas de IA?

Algunos poderes legislativos, como el de los Estados Unidos, están intentando regular temas tales como: a) Algoritmos de *Machine Learning – Deep Learning*; b) Rastreo del error y medición del daño; c) Responsabilidad; d) Opacidad; e) Privacidad; f) Explicabilidad; g) Interpretabilidad; h) Auditabilidad; i) Riesgos; y, j) Derechos Humanos.¹⁶ ¿Qué estamos haciendo en México? por el momento nada similar. Esto nos lleva nuevamente a la necesidad de generar esta discusión dentro

¹⁶ Dos proyectos de legislación que son recientes, los encontramos en los Estados Unidos. “El primero pretende modificar el código administrativo de la ciudad de Nueva York, con el fin de regular la venta de herramientas automatizadas de decisión utilizadas para la selección de personal. Su iniciativa está actualmente en el proceso de discusión legislativa. El segundo tiene que ver con toda una ley de Responsabilidad Algorítmica de 2019, la cual busca que las empresas que utilizan sistemas de inteligencia artificial, cuando involucren información personal o tomen decisiones automatizadas, realicen evaluaciones de sus algoritmos. Específicamente, los sistemas de decisión automatizados “de alto riesgo” como por ejemplo aquellos que pueden contribuir a la inexactitud, el sesgo o la discriminación; o aquellos que pueden facilitar la toma de decisiones sobre aspectos sensibles de la vida de las personas mediante la evaluación de su comportamiento.”. Cfr. José Enrique Siqueiros Fernández y José Luis Hernández Sánchez, *Ingeniería, Filosofía y Derecho, Diálogo hacia una perspectiva ecosistémica de la Ética en Inteligencia artificial*, s/f, s/e [documento en revisión para su publicación por la Universidad Panamericana, Facultad de Filosofía, México]

del Poder Legislativo, con el fin de construir el andamiaje regulatorio que fomente el uso de la IA con bases éticas y jurídicas.

La **metodología** utilizada en esta investigación se basa en dos libros. Del primero, rescata los consejos generales que todo estudiante debe seguir para la redacción de un trabajo doctoral y fue escrito por Jaime Nubiola: *El taller de la Filosofía. Una introducción a la escritura filosófica*¹⁷. El segundo, da soporte a toda la estructura conceptual y argumentativa de cada uno de los apartados de esta tesis, la cual busca un discurso coherente y bien organizado (del que precisamente habla Nubiola), que sea claro, basado en la lógica y en la objetividad, que vaya más allá de lo descriptivo, de vanguardia y que busque hacer aportaciones a la ciencia jurídica y, quizás, a otras disciplinas. El libro que utilizamos se intitula: *Metodología de la Disertación Filosófica y Jurídica* del abogado y doctor por la Universidad de Buenos Aires, Germán Sucar¹⁸.

Cabe señalar que alcanzar la unión de los ingredientes anteriores no es tarea sencilla. Mucho menos cuando el perfil del doctorando proviene de una formación profesional vinculada más a la administración pública y a la empresa que al mundo de la filosofía y la investigación jurídica. Esto tiene más desventajas que oportunidades para la construcción de un trabajo doctoral, sobre todo si se pretende utilizar la metodología americana de las ciencias sociales o la metodología de investigación jurídica que la mayoría de las Facultades de Derecho del país contempla para estos fines. Y me atrevo a afirmar que esto influye en el doctorando para llegar a la conclusión de un trabajo doctoral.

A diferencia de los estudiantes de doctorado de otras latitudes (o de algunas minorías en México), que se encuentran de tiempo completo elaborando investigación y la combinan con espacios para actividades docentes o de corte académicas que los llevan de manera inercial a su incorporación como parte del claustro de alguna universidad, en México esta no es la regla general y gran parte de los estudiantes de doctorado van más hacia perfiles de profesionistas que se han formado fuera de aula y los talleres de investigación jurídica.

¹⁷ Jaime Nubiola, *El taller de la Filosofía. Una introducción a la escritura filosófica*, Pamplona, Ediciones Universidad de Navarra, 2017.

¹⁸ Germán Sucar, *Metodología de la disertación filosófica y jurídica*, Valencia, Tirant lo Blanch, 2019.

Debo mencionar que desde el año 2000 he venido combinando el ejercicio profesional en el ámbito gubernamental y de empresa privada, con espacios dedicados a mis estudios de posgrado y que durante estos años he indagado sobre el uso de diversas metodologías para la investigación jurídica¹⁹ y considero que la metodología utilizada por Sucar ofrece resultados más efectivos para quienes combinan sus actividades profesionales con la redacción de una tesis doctoral.

Sucar explica que, en principio, el estudiante debe de hacer una lectura intensa sobre el tema de interés que quiere investigar y organizar la bibliografía consultada siguiendo un plan riguroso de trabajo para la construcción de una introducción y la categorización de la información encontrada en dos partes, cada una dividida a su vez en dos subpartes (que en suma proyectan el fondo o contenido y la forma o estructura), esto es algo bastante complejo de conseguir, ya que requiere alta capacidad de análisis y de síntesis, así como un ejercicio de lecturas previas bastante ambicioso, pero una vez que se logra la articulación de las 2 partes y sus subpartes, se avanza con mayor rapidez y los resultados van apareciendo de la mano de la revisión del propio texto, corrigiendo y relaborando las fallas con la propia bibliografía consultada.

Reconocimiento especial merece el doctor Jorge Cerdio Herrán, director de este trabajo doctoral, ya que siempre estuvo al pendiente de las lecturas y de la construcción de la estructura comentada. Agradezco la confianza brindada y sobre todo la posibilidad de permitirme construir ideas en temas tan complejos. Aclaro también que cualquier error o mal planteamiento conceptual es responsabilidad del

¹⁹ Entre las obras clásicas de metodología que previamente consultamos para evaluar y decidir la metodología a seguir en nuestra investigación están: Miguel Villoro Toranzo, *Metodología del trabajo jurídico: Técnicas del Seminario de Derecho*, México, Limusa, 1999; Ignacio Méndez Ramírez, *et. al.*, *El protocolo de investigación. Lineamientos para su elaboración y análisis*, México, Trillas, 2004; Heinz Dieterich Steffan, *Nueva Guía para la investigación científica*, Lima, Universidad de Ciencias y Humanidades, 2008; Héctor Fix-Zamudio, *Metodología, Docencia e Investigación Jurídicas*, México, Porrúa, 2007; Umberto Eco, *Cómo se hace una tesis*, Barcelona, Gedisa, 2010; Roberto Hernández-Sampieri, Sergio Méndez Valencia, *et. al.*, *Fundamentos de investigación*, Mc Graw Hill, México, 2017; Roberto Hernández-Sampieri y Christian Paulina Mendoza Torres, *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*, México, Mc Graw Hill, 2019; Luis Ponce de León Armenta, *Metodología del derecho*, México, Porrúa, 2020.

que escribe estas líneas y no del director, quien siempre ha sido respetuoso de mis planteamientos.²⁰

Abundando más sobre la metodología, debo decir que los beneficios que fue generando el libro de Sucar también los encontramos en el desarrollo de hábitos para una lectura y escritura funcional y, como dice el autor, utilizar “las pautas que ofrece para el adecuado tratamiento, estructuración y presentación de diversos trabajos teóricos, especialmente en el ámbito de la filosofía y del derecho”²¹. Este binomio no lo contemplan otras metodologías y da fortaleza al estudiante de doctorado y a su trabajo de tesis.

¿De dónde toma Sucar esta metodología? Sucar deja a un lado el modelo de metodología anglosajona y toma como base el modelo de metodología francesa, bajo una premisa de gran relevancia que omiten otros autores de libros similares: reconocer que existe una falta de habilidades de lectura y escritura jurídicas de los alumnos.²² Esto es un problema del estudiante en muchas partes del mundo, que

²⁰ El doctor Jorge Cerdio Herrán, es profesor de Derecho e Informática del Doctorado en Derecho de la Universidad Panamericana, director del Centro de Acceso a la Justicia del ITAM, miembro del *Policy Development Group* del CAHAI-COE (Unión Europea) y ex director de la Facultad de Derecho del Instituto Tecnológico Autónomo de México. Para conocer algunas aportaciones recientes del doctor Cerdio, relacionadas con el uso de sistemas inteligentes, protección de Derechos Humanos y la relevancia de la filosofía como instrumento para la legislación que se está construyendo en la Unión Europea, se pueden consultar dos conferencias: Una, ante la Fletcher School (9 de marzo de 2021) en la que se explica cómo esta revolución de la IA nos lleva, inevitablemente, a recurrir al régimen internacional de Derechos Humanos como el paradigma legal e institucional a través del cual debemos evaluar estos cambios, sus impactos y las posibles soluciones. Y, sobre todo, el doctor Cerdio da la pauta para entender la importancia de la discusión enfocada a las cuestiones de cómo evaluar el impacto de la IA en los Derechos Humanos y cómo los Derechos Humanos pueden moldear y regular el desarrollo de la IA. Consúltese: *Human Rights in the Era of Artificial Intelligence*: <https://www.youtube.com/watch?v=zkl1W9CcdxM> [20 de julio de 2022] Y también su participación en el Primer Congreso de Lógica, Epistemología y Ética de la Inteligencia Artificial organizado por la Facultad de Filosofía y la Facultad de Ingeniería de la Universidad Panamericana en donde expuso el tema: “*Human Rights, Democracy and Rule of Law Impact Assessment of AI systems*”, el Proyecto de la Unión Europea, en el que expone los principales cambios legislativos que están por aprobarse para regular la IA en la Unión Europea y que, inevitablemente, marcarán el rumbo de las políticas regulatorias nacionales y supranacionales en muchas partes del mundo. Consúltese: <https://fb.watch/v/2VFm7bJet/> [20 de julio de 2022]

²¹ *Idem*, p. 19.

²² Dicho tipo de metodología y de enseñanza, está ausente en las universidades mexicanas. Sucar trabajó 10 años en el diseño de la obra y nos explica su razón de ser, siguiendo a Gilles Boveaux y Philippe Bihr: “el objeto de la disertación [jurídica] es la exposición, la explicación, la discusión, de las reglas de derecho relativas a una cuestión determinadas. El deber ser organizado según un plan riguroso que tradicionalmente contiene una introducción bastante amplia y dos partes, ellas mismas subdivididas cada una en dos subpartes, mientras que la conclusión no es generalmente necesaria”. Y continúa diciendo: “Así hay dos grandes aspectos que hay que tener en cuenta al proyectar una

a veces no se quiere reconocer. Sin embargo, debemos aceptar que en los estudios de posgrado hay esa carencia, y también, que existen soluciones metodológicas como la que se utiliza en esta investigación ya que fomenta el hábito de estudio mediante el desarrollo de la capacidad de análisis y la capacidad de síntesis.

La realidad es que muchos estudiantes de doctorado tienen deficiencias. En el ámbito jurídico, existen grandes abogados, jueces, fiscales, empresarios que son alumnos del doctorado y ven sesgada la conclusión de esta etapa por que el modelo de metodología anglosajona busca una articulación del trabajo doctoral más compleja y menos flexible para este tipo de perfil profesional. Debemos entender también que es lo mismo redactar una demanda, un oficio, un ensayo o artículo en prensa que elaborar una disertación científica (ya sea filosófica o jurídica) y Sucar pone ejercicios prácticos para llegar a este objetivo.

Estamos conscientes de que esta metodología no es la más conocida y que pudiera generar rechazo (ya que la tendencia es acudir a obras más tradicionales). Sin embargo, decidimos mantenerla como un ingrediente novedoso de este trabajo doctoral, invitando a los profesores y alumnos a utilizarla y comprobar los beneficios de la metodología francesa; sobre todo en temas que requieren una visión transdisciplinaria.

Lo novedoso de esta investigación va también en el uso de una metodología alterna, que permitiera abrir la mente del abogado a la experimentación científica en Inteligencia Artificial, en el uso de la Ética y el Derecho. Todo se unió en un espacio propicio para aplicar estas novedades metodológicas. Máxime que la Inteligencia Artificial y sus aportaciones requieren de un análisis transdisciplinario que sale de los parámetros establecidos en las metodologías jurídicas convencionales. Esta limitante también se puede apreciar en la Sociología del Derecho o en la Antropología Jurídica, disciplinas en las que método jurídico fue pensado para un rigor académico propio, basado en la norma y el hecho social, pero sin posibilidad

disertación jurídica: el fondo y la forma. En cuanto al primero, baste con decir aquí que su objeto es la exposición, explicación y discusión de normas jurídicas relativas a una cuestión determinada. La aproximación al tema es, pues, "teórica" pero ello no en un sentido que no deba tenerse en cuenta las aplicaciones prácticas concernidas... En cuanto a la forma, tal como se indica en los apartados siguientes, se trata, en síntesis, de ordenar los desarrollos según un plan muy estricto." Germán Sucar, *Metodología de la disertación filosófica y jurídica*, Valencia, Tirant lo Blanch, 2019.

de generar un conocimiento transdisciplinario para la construcción de su hipótesis, planteamientos del problema, objetivos y conclusiones.

Por otro lado, se aclara que la disertación jurídica que se hace siguiendo la metodología francesa tiene **tres pautas argumentativas**. La primera está en la **exposición** de la evolución de la IA y sus principales conceptos para llegar a lo que se conoce como Sistemas Expertos Decisionales, Machine Learning y Deep Learning. Estos son los dos tipos de IA más utilizados para la construcción de sistemas autónomos que auxilian a la toma de decisiones, de aquí también aparece la computación cognitiva y otros modelos de IA basados más en la biología que en la lógica tradicional. Una vez explicado esto, se desarrollan tanto los problemas de funcionamiento como la rendición de cuentas. Dos temas ampliamente discutidos por la doctrina y que abarcan una temática diversa (datos, sesgos, opacidad, interpretabilidad, explicabilidad) que conduce la investigación hacia la búsqueda de la transparencia, de la privacidad y de la auditabilidad algorítmica.

La segunda pauta argumentativa la encontramos al hacer una **explicación** de las aportaciones que ha hecho IBM y el sistema Watson para llegar a su vinculación con las ciencias de la salud, el acto médico y el uso de *Watson Health*. Para ello y ante la escasez de información que existe en el foro jurídico sobre este tema, nos dimos a la tarea de vincularnos con textos de autores de múltiples disciplinas, lo cual permitió ampliar nuestro panorama de conocimientos para ir correlacionando conceptos y creando argumentos de forma transdisciplinaria para que el experto en la rama del derecho (abogados, jueces, fiscales, peritos, legisladores, académicos), conozca los avances de la Inteligencia Artificial desde su raíz y las novedades más relevantes hasta ahora.

Esta segunda pauta argumentativa no hubiera sido posible sin la guía invaluable del doctor Jorge Cerdio Herrán y sin la asistencia a diversos cursos sobre el uso de la IA. En este orden de ideas, el funcionamiento del sistema Watson, sus servicios y el sistema *Watson Health* fueron estudiados a través de los *Massive Open Online Course (MOOC's)* Coursera y Edx a través de tres cursos de capacitación que ofrece IBM: 1) *Introduction to Artificial Intelligence*, 2) *Getting*

Started with AI using IBM Watson y 3) Introducción a la ciencia de datos y sus aplicaciones.

En cuanto a fuentes de información relativas a los algoritmos de *Machine Learning – Deep Learning*, el trabajo se nutrió de la Biblioteca digital ACM (<https://dl.acm.org/>), una plataforma de investigación que contiene el texto completo de muchas de las publicaciones de la *Association for Computing Machinery* (ACM) y de otras editoriales y una base de datos muy completa sobre el campo de la informática. La ACM fue creada en 1947, reúne a educadores, investigadores y profesionales de la informática de todo el mundo; es considerada la sociedad informática más grande del planeta (más de 100 mil miembros). Estos sitios cuentan con gran cantidad de información sobre el uso de la Ética en la IA que no es utilizada por abogados y que requiere su uso para ampliar conceptos y novedades al sistema jurídico.

Finalmente, la tercera pauta argumentativa está en la **discusión** de las categorías de responsabilidad jurídica, el uso de la Ética empresarial, así como de la propuesta de incorporación de conceptos nuevos tales como: sesgo algorítmico, responsabilidad algorítmica, justicia algorítmica y agente moral artificial autónomo.

Adicionalmente a lo anterior, se desarrolla el **concepto responsabilidad desde tres distintas dimensiones** que permiten medir los riesgos legales, tecnológicos y éticos en una tabla de indicadores que pueden ser incorporados por las áreas de *Compliance*. Las dos primeras dimensiones, son ontológicamente idénticas, pero epistémicamente distintas; y la tercera, tiene que ver con el papel de la Ética empresarial y el uso de la IA.

En la primera dimensión, encontramos a la norma vigente y la dogmática que le da vida y que le permite enjuiciar por un hecho ilícito o por la comisión de un delito a cualquier persona que esté en los supuestos contenidos en la propia legislación. Aquí surge la interrogante ¿cómo funcionan actualmente las categorías vigentes de responsabilidad jurídica y su dogmática para saber si pueden solucionar los problemas que surgen por el uso de sistemas de IA como lo es *Watson Health*? y

se analiza qué salidas ofrece la legislación ante un daño o un impacto adverso a Derechos Humanos por el uso de la tecnología.

En la segunda dimensión, la que aún no está vigente en nuestro sistema jurídico ni en muchos otros, se analizan nuevos conceptos como el de tipo de inteligencia artificial, datos, opacidad, sesgos algorítmicos, interpretabilidad, explicabilidad, y la posibilidad de utilizarlos para construir el concepto de Agente Artificial Moral Autónomo (AAMA), que a su vez permiten dividir la responsabilidad que le corresponde a la máquina (responsabilidad algorítmica) de la responsabilidad que le corresponde a los seres humanos, como la de los que producen los datos, los que procesan los datos, los que desarrollan los sistemas; los que utilizan el sistema; los que adquieren el sistema; los que certifican un sistema; y los que auditan el sistema.

Los legisladores, jueces, fiscales, peritos, abogados, científicos de datos, empresas de IA; y todas las personas que estén vinculadas al ámbito del Derecho y la IA, deben entender estos nuevos conceptos e incluirse dentro de las áreas de *Compliance* como indicadores de riesgos por el uso de IA. Así como hubo una actualización al concepto responsabilidad penal de las personas jurídicas, hoy debemos hacer otra actualización conceptual relacionada con el uso de determinados modelos de IA que pueden generar daños o impactos adverso a los Derechos Humanos de las personas.

Y, finalmente, la tercera dimensión tiene que ver con algunas aportaciones de la Ética empresarial, que para el caso concreto utilizamos las aportaciones de Carlos Llano Cifuentes. Aquí la pregunta es ¿si a la par de una respuesta jurídica debería haber una respuesta Ética? Nosotros creemos que sí y que las aportaciones de la Ética empresarial de Carlos Llano Cifuentes dan el soporte adecuado para ello; sin que esto sea obstáculo para que las empresas o los expertos, utilicen otra corriente filosófica para la medición de sus riesgos éticos.

El autor de este texto está convencido de que Carlos Llano lleva la Filosofía al mundo empresarial y construye lo que llamamos la Teoría de una Moral de Responsabilidad Empresarial con sus cuatro campos de actuación: la

responsabilidad del consecuente (o responsabilidad de las consecuencias de la acción), la responsabilidad del antecedente (o responsabilidad de los principios), la responsabilidad congruente (o responsabilidad de los proyectos anteriores de mi vida); y la responsabilidad del trascendente (o de la misión a la que, con mi vida, he sido destinado). La importancia de utilizar esta aportación como mecanismo de autorregulación o de integridad empresarial, permite complementar el uso del *Compliance*. La empresa, por tanto, tendrá elementos que demuestren a la autoridad que hizo todo lo que estuvo en sus manos para evitar una afectación o impacto adverso a Derechos Humanos, y con ello, evitar una posible responsabilidad jurídica (o quizás, algorítmica).

Llano conocía perfectamente las distintas posturas de los abogados en el tema de la responsabilidad (Kant, Rawls, Dworkin, entre otros) y con su pensamiento las analizó y decidió construir categorías de responsabilidades específicas para la empresa. La empresa para Llano no debe ser entendida como un lugar de comercio de productos o servicios, sino como una comunidad de seres humanos con dignidad, inteligencia, fines en común y responsabilidades. No en vano se le conoce a Carlos Llano Cifuentes como el filósofo-empresario y el empresario-filósofo.

Diremos finalmente, que la metodología que propone Sucar se inspira en diversas obras francesas de probada eficacia metodológica, encontrando evidencias de su uso tanto en autores clásicos del Derecho como Henri Capitant (y su obra *Comment il faut faire sa thèse de doctorat en droit*) como en autores más contemporáneos tales como Édith Jaillardon y Dominique Roussillon (*Outils pour la recherche juridique. Méthodologie de la thèse de doctorat et du mémoire de master en droit*)²³. Tanto en uno como en los otros, encontraremos un camino alternativo para la elaboración de un trabajo doctoral

²³ *Ibidem*, p. 528. En Francia se exige que las tesis doctorales no solamente demuestren dominio de una especialidad sino también la cultura científica del candidato. Cfr. Arístides Alfredo Vara Horna, ¿Cómo evaluar la rigurosidad científica de las tesis doctorales? Criterios teóricos y metodológicos aplicados a la educación, Perú, Universidad de San Martín de Porres, 2010, p. 64 y ss.

“Disertar”, nos comenta Sucar, es “razonar, discurrir detenida y metódicamente sobre alguna materia, bien para exponerla, bien para refutar opiniones ajenas...”²⁴ en la disertación jurídica es su autor quien deberá, **sobre la base de determinada bibliografía consultada (universo bibliográfico), delimitar el tema, plantear el problema, enunciar la tesis, articular una problemática y desarrollar dialécticamente los argumentos correspondientes... Uno de los objetivos o finalidades centrales de este ejercicio es favorecer la elaboración de trabajos problemáticos y argumentativo-dialécticos... o, dicho de otro modo, evitar los trabajos puramente descriptivos...**²⁵

En este sentido la *Metodología de la Disertación Filosófica y Jurídica* no permitió dar a este trabajo dos características importantes que se deben tomar en cuenta para una disertación jurídica doctoral: La primera se relaciona con cuestiones de fondo relacionadas a la exposición, explicación y discusión de normas jurídicas soportadas en la Teoría de la Responsabilidad Jurídica (y su soporte moral), que para el caso particular la encontramos en la búsqueda y análisis de las normas nacionales e internacionales (tanto de *soft law* como de *hard law*).

La segunda característica se encuentra relacionada con la estructura y se trata de saber ordenar amplias cantidades de información como argumentos siguiendo “un plan binario muy estricto” que te permite categorizar argumentos y descartar paralogismos, que por cierto, Sucar retoma de Gobeaux y Bihr²⁶ bajo sus conceptos: títulos, *chapeaux* (encabezamientos) y transiciones. En conjunto, permiten que la disertación del doctorando hable con mayor soporte científico, expresando ideas articuladas en su estructura.²⁷

Gracias a esta metodología que empleamos se pudieron estudiar conceptos diversos al ámbito jurídico (Inteligencia Artificial basada en Sistemas Expertos, Inteligencia Artificial basada en Machine Learning y Deep Learning), pero bajo una hipótesis y un planteamiento del problema derivado de normas jurídicas.

²⁴ *Ibidem*, p. 24.

²⁵ *Ibidem*, p. 245.

²⁶ *Idem*.

²⁷ *Idem*.

Una vez hecho lo anterior, se hace un ejercicio dialéctico que analiza los problemas de funcionamiento de la segunda tecnología (Machine Learning – Deep Learning), lo que trae como resultado la necesidad de examinar la rendición de cuentas por el uso de la Inteligencia Artificial (tanto dentro de las empresas, como en los gobiernos y la sociedad en conjunto). Con estos dos eslabones conceptuales se lleva la investigación hacia el caso Watson Health, describiendo lo que es IBM Watson y el caso Watson Health. Y finalmente, se vincula todo lo anterior, con la dimensión de la responsabilidad jurídica y su aplicación a casos de sistemas de IA para la toma de decisiones, así como el uso de la Ética empresarial (con el catálogo de indicadores para medir riesgos legales, tecnológicos y éticos). Esto no lo permite la metodología jurídica tradicional.

La *Metodología de la Disertación Filosófica y Jurídica* nos dio las pautas “para un adecuado tratamiento, estructuración y presentación de ciertos trabajos teóricos, especialmente en el ámbito de la filosofía y del derecho y, más específicamente, en la redacción de géneros jurídicos con una impronta filosófica.”²⁸. También permitió un “adecuado equilibrio entre contenido, estructura y estilo, entre análisis y síntesis, de instar a la reflexión crítica; de desarrollar una argumentación consistente con precisión, coherencia y claridad”²⁹.

Finalmente, quiero agradecer al doctor Dr. Víctor-Isolino Doval, profesor investigador en el Instituto de Humanidades de la Universidad Panamericana, los comentarios hechos para fortalecer la metodología y las reflexiones sobre Carlos Llano Cifuentes. Y de manera sempiterna, mi agradecimiento al doctor Hugo Saúl Ramírez García, director del Doctorado en Derecho de Facultad de Derecho de la Universidad Panamericana, quien conformó muy atinadamente un sínodo con perfiles transdisciplinarios que era necesario para el análisis y dictaminación (a doble ciego) de esta tesis. Este es un ingrediente muy útil para incrementar la calidad de las tesis doctorales en la Facultad de Derecho de la Universidad Panamericana.

²⁸ German Sucar, op. cit., p. 19.

²⁹ *Idem*.

Capítulo I. Inteligencia Artificial

Capítulo I. Inteligencia Artificial

En este apartado se propone como **objetivo particular**, exponer la evolución, conceptos, tipos y principales características de la Inteligencia Artificial basado en Sistemas Expertos Decisionales **(A)** y la Inteligencia Artificial basada en Machine Learning – Deep Learning **(B)**. Aclaremos desde ahora que la mayor parte de los sistemas de IA son controlados por voluntad del ser humano, no ocasionan daños y no afectan Derechos Humanos; prácticamente, el funcionamiento de sus algoritmos y sus resultados son predecibles y controlables (de inicio a fin). Pero se ha comprobado que existe un tipo de IA (la del inciso **B**) que sola o combinada con la del inciso **(A)** puede llegar a generar algún daño o la a Derechos Humanos. En este tipo de IA (que es el inicio de la llamada IA generalizada) no está presente la voluntad del ser humano o, por lo menos, no queda clara la participación de personas en las decisiones que efectúan los sistemas autónomos para la toma de decisiones. Esa mínima parte, vista como la que origina las afectaciones a Derechos Humanos, es la que requiere estudio, investigación y atención transdisciplinaria por los ingenieros, filósofos y abogados.

Dicho lo anterior, lo primero que se debe entender es cómo surge el concepto IA, algunas definiciones y los tipos de IA que auxilian al hombre para tomar decisiones (Sistemas Expertos Decisionales, *Machine Learning – Deep Learning*). Vale la pena aclarar que las reflexiones que se harán a lo largo de este trabajo giran en torno a aquellos sistemas de IA que auxilian al ser humano a la toma de decisiones, es decir, se debe diferenciar entre la IA que auxilia al ser humano en la toma de decisiones y no otros tipos de IA, como la robótica o la visión computacional cuyo estudio y consecuencias en el ámbito de la responsabilidad jurídica cambia. La primera tecnología tiene una realidad virtual, mientras que la segunda tiene una realidad física.³⁰ Un ejemplo claro de lo anterior, lo encontramos con la delimitación que hacen Stuart Russell y Peter Norving de la IA en seis categorías: 1.

³⁰ Agradezco al doctor Hiram Ponce, director del Doctorado en Inteligencia Artificial de la Universidad Panamericana, esta última reflexión, ya que es frecuente que la comunidad jurídica -y quizás otras profesiones- entremezcle este tipo de conceptos sin entender las consecuencias que pudieran generar para ubicar responsabilidad jurídica de las personas y la reparación del daño a las víctimas.

Procesamiento de lenguaje natural; 2. Representación del conocimiento; 3. Razonamiento automático; 4. Aprendizaje automático; 5. Visión computacional; y, 6. Robótica. A la vez, definen a la Inteligencia artificial como “el estudio de los agentes que reciben percepciones del entorno y llevan a cabo las acciones. Cada agente implementa una función la cual estructura las secuencias de las percepciones en acciones; también tratamos las diferentes formas de representar estas funciones, tales como sistemas de producción, agentes reactivos, planificadores condicionales en tiempo real, redes neurales y sistemas teóricos para las decisiones.”³¹

¿Cómo surge el concepto IA? La conferencia sobre ciencias de la computación en el Dartmouth College de 1956, en *New Hampshire*, Estados Unidos es considerada por los especialistas como el momento histórico en el que se formaliza el uso del término la inteligencia artificial. Los temas que se discutieron en esa ocasión fueron: “1. Automatic Computers; 2. How Can a computer be Programmed to Use a Language; 3. Neuron Nets; 4. Theory of the Size of a Calculation; 5. Self-Improvement; 6. Abstractions; y 7. Randomness and Creativity”³². Desde esta fecha podemos encontrar discusiones sobre los distintos tipos de IA, su uso, el interés de los matemáticos por lograr un lenguaje de comunicación hombre-máquina, así como el uso de la lógica y la biología para entender el cerebro humano, su inteligencia y la forma en cómo toma decisiones.

El concepto de IA puede ser considerado multívoco y ha venido cambiando, dependiendo el lugar, la época y la corriente de investigación. Para efectos de acotar la discusión de nuestra investigación hemos seleccionado algunos conceptos que consideramos clave para entender este tema.

³¹ Cfr. Inteligencia Artificial. Un enfoque moderno, segunda edición, Pearson Educación, Madrid, 2004, pp. XIX, 37 a 62. Un agente -dicen los autores- “es cualquier cosa capaz de percibir su medioambiente con la ayuda de sensores y actuar en ese medio utilizando actuadores (se usa este término para indicar el elemento que reacciona a un estímulo realizando una acción. Un agente humano tiene ojos, oídos y otros órganos sensoriales además de manos, piernas, boca y otras partes del cuerpo para actuar. Un agente robot recibe pulsaciones del teclado, archivos de información y paquetes vía red a modo de entradas sensoriales y actúa sobre el medio con mensajes en el monitor. El término percepción se utiliza en este contexto para indicar que el agente puede recibir entradas en cualquier instante. Un agente racional es aquel que hace lo correcto; en términos conceptuales...”

³² En el Anexo 1 de este apartado se pueden leer los detalles de cada tema.

Uno de los conceptos de inteligencia artificial más citado es el de Marvin Minsky: “La inteligencia artificial es la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hicieran los humanos”.³³ Otra definición de inteligencia artificial que es muy respetada es la del IEEE, que la define como la “capacidad de las computadoras o de otras máquinas de mostrar o simular inteligencia natural”.³⁴

Hoy en día, como diría Lasse Rouhiainen, la IA debe ser entendida como “... la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones, tal y como lo haría un ser humano”.³⁵

La Comisión Europea, creó el Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial (junio 2018) el cual publicó en 2019 una definición de Inteligencia Artificial o, mejor dicho, de “Sistema de Inteligencia Artificial” que está siendo utilizada por varias autoridades en la construcción de normas en la Unión Europea. El llamado “Libro Blanco sobre la inteligencia artificial -un enfoque europeo orientado a la excelencia y la confianza” retoma la definición construida por el Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial: “Los sistemas de inteligencia artificial (IA) son programas informáticos (y posiblemente también equipos informáticos) diseñados por seres humanos que dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno mediante la adquisición de datos, la interpretación de los datos estructurados, el razonamiento sobre el conocimiento o el tratamiento de la información, fruto de estos datos y las mejores acciones que se llevarán a cabo para alcanzar el objetivo fijado.”.³⁶

³³ Para más detalles sobre este tema, consúltase el Anexo 1 de esta sección. Cfr. Marvin Minsky, *Semantic Information Processing*, MIT Press, 1968, p. V.

³⁴ Standards Committee of the IEEE Systems, Man, and Cybernetics Society, *IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being*, IEEE SA Standards Board, 2020, p. 18.

³⁵ Rouhiainen, Lasse, *Artificial Intelligence: 101 Things You Must Know Today About Our Future - Updated Edition for Post-Covid-19 World*. Edición de Kindle, 2020.

³⁶ Comisión Europea, *Libro Blanco sobre la inteligencia artificial -un enfoque europeo orientado a la excelencia y la confianza*, Bruselas, 2020, p. 20. Esta definición tiene un antecedente previo en la Comunicación de la Comisión, de 25 de abril de 2018, titulada “Inteligencia Artificial para Europa” (COM(2018)0237) que señala: “El término “inteligencia artificial” (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la

En 2019, el Congreso y el presidente de los Estados Unidos, se dieron a la tarea de crear la Comisión de Seguridad Nacional sobre Inteligencia Artificial (*National Security Commission on Artificial Intelligence, NSCAI*). En la Comisión participan 15 expertos en tecnología, ejecutivos de empresa, académicos y profesionistas líderes en tecnología y seguridad nacional (comisionados), designados por consenso bipartidista y con la anuencia del Poder Ejecutivo para avanzar -dice su texto- en el desarrollo de la Inteligencia Artificial, el aprendizaje automático y las tecnologías asociadas para abordar de manera integral las necesidades de defensa y seguridad nacional de la Estados Unidos. La comisión está facultada para examinar temas de competitividad nacional, analizar las ventajas tecnológicas de Estados Unidos en el mundo, realizar acciones de cooperación internacional y de competitividad internacional, fomentar inversión en investigación básica y avanzada, fuerza laboral y capacitación, riesgos potenciales del uso militar, preocupaciones Éticas, establecimiento de estándares de datos y la evolución futura de la IA.³⁷

acción -con cierto grado de autonomía- con el fin de alcanzar objetivos específicos. Los sistemas basados en IA pueden consistir simplemente en un programa informático (p. ej. Asistentes de voz; programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. Robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas)". Consúltese también las tres nociones de inteligencia artificial (No técnica; técnica y de política pública o regulatoria) en: Parlamento Europeo, Policy Department for Citizens, Rights and Constitutional Affairs, Committee on Legal Affairs, *Artificial Intelligence and Civil Liability*, Legal Affairs, Parlamento Europeo, Bruselas, 2020. Consúltese en: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU\(2020\)621926_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf) [20 de julio de 2022]

³⁷ La Ley de Autorización de Defensa Nacional John S. McCain para el año fiscal 2019 incluye la siguiente definición de IA para guiar el trabajo de la Comisión: 1. Cualquier sistema artificial que realice tareas en circunstancias variables e impredecibles sin supervisión humana significativa, o que pueda aprender de la experiencia y mejorar el desempeño, cuando se expone a conjuntos de datos. 2. Un sistema artificial desarrollado en software de computadora, hardware físico u otro contexto que resuelve tareas que requieren percepción, cognición, planificación, aprendizaje, comunicación o acción física similares a las de los humanos. 3. Un sistema artificial diseñado para pensar o actuar como un ser humano, incluidas arquitecturas cognitivas y redes neuronales. 4. Un conjunto de técnicas, incluido el aprendizaje automático, diseñado para aproximarse a una tarea cognitiva. 5. Un sistema artificial diseñado para actuar racionalmente, que incluye un agente de software inteligente o un robot incorporado que logra objetivos mediante la percepción, la planificación, el razonamiento, el aprendizaje, la comunicación, la toma de decisiones y la actuación.

La *National Security Commission on Artificial Intelligence*, retoma una definición de IA del Director de Google Drive, Andrew Moore³⁸ y otros dos expertos³⁹ no es una sola pieza de hardware o software, sino más bien una constelación de tecnologías que dependen de elementos interrelacionados que pueden concebirse como una pila, que requieren percepción, cognición, planificación, aprendizaje, comunicación o acción física similares a las humanas; y tecnologías que pueden aprender y actuar de forma autónoma, ya sea en forma de agentes de software o robots incorporados.⁴⁰

³⁸ El Dr. Andrew W. Moore es un científico informático con experiencia en aprendizaje automático y robótica. Se convirtió en el jefe de la división de Inteligencia Artificial de Google Cloud en enero de 2019. Moore trabajó anteriormente en Google de 2006 a 2014 y fue el director fundador de la oficina de Ingeniería de Pittsburgh de Google en 2006. Los intereses de investigación de Moore abarcan el campo de los “macrodatos”: la aplicación de métodos estadísticos y fórmulas matemáticas a cantidades masivas de información, que van desde búsquedas en la web hasta astronomía y registros médicos, con el fin de identificar patrones y extraer significado de esa información. Su investigación anterior ha incluido la mejora de la capacidad de los robots y otros sistemas automatizados para detectar el mundo que los rodea y responder de manera adecuada.

³⁹ Cfr. Andrew W. Moore, Martial Hebert, Shane Shaneman, “The AI stack: a blueprint for developing and deploying artificial intelligence”, *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent*, Proc. SPIE 10635, ISR IX, 106350C (*National Security* 4 de mayo de 2018); <https://doi.org/10.1117/12.2309483> [20 de julio de 2022]

⁴⁰ Cfr. National Security Commission on Artificial Intelligence (NSCAI), *The NSCAI Final Report*, 2021. Consúltese en: <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf> [20 de julio de 2022]

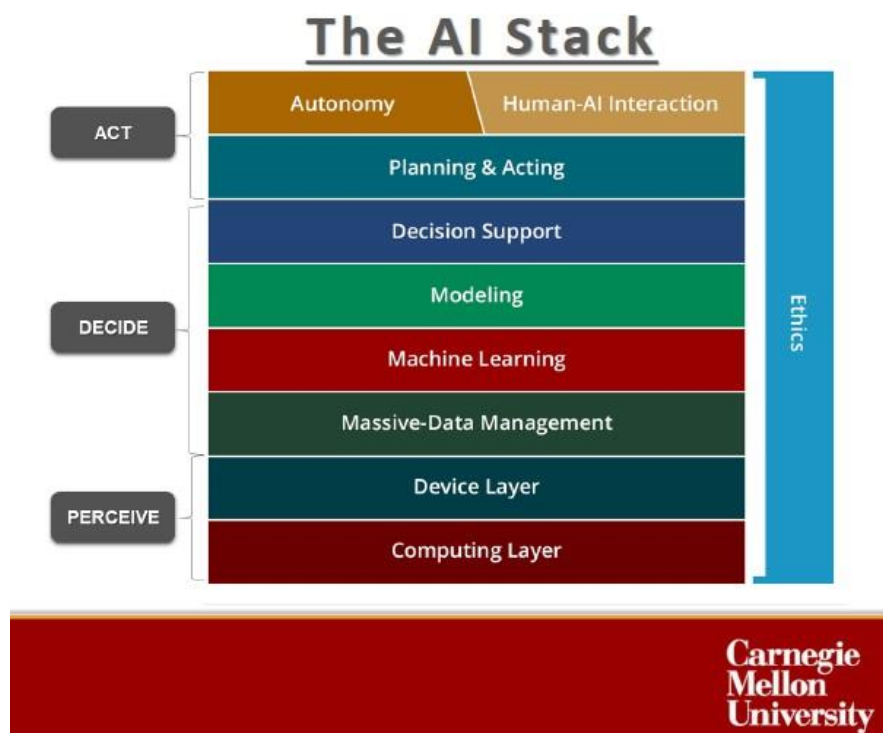
Tabla de definiciones sobre “Inteligencia Artificial” o “Sistemas de Inteligencia Artificial”

Autor	Definición
Marvin Minsky	<i>...es la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hicieran los humanos</i>
Stuart Russell y Peter Norving	<i>...el estudio de los agentes que reciben percepciones del entorno y llevan a cabo las acciones. Cada agente implementa una función la cual estructura las secuencias de las percepciones en acciones; también tratamos las diferentes formas de representar estas funciones, tales como sistemas de producción, agentes reactivos, planificadores condicionales en tiempo real, redes neurales y sistemas teóricos para las decisiones</i>
Lasse Rouhiainen	<i>...la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones, tal y como lo haría un ser humano</i>
Comisión Europea ⁴¹	<i>...se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción -con cierto grado de autonomía- con el fin de alcanzar objetivos específicos... que luego evoluciona para decir: “Los sistemas de inteligencia artificial (IA) son programas informáticos (y posiblemente también equipos informáticos) diseñados por seres humanos que dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno mediante la adquisición de datos, la interpretación de los datos estructurados, el razonamiento sobre el conocimiento o el tratamiento de la información, fruto de estos datos y las mejores acciones que se llevarán a cabo para alcanzar el objetivo fijado.”.</i>
National Security Commission on Artificial Intelligence (Andrew W. Moore, Martial Hebert, Shane Shaneman)	<i>IA no es una sola pieza de hardware o software, sino más bien una constelación de tecnologías que dependen de elementos interrelacionados que pueden concebirse como una pila, que requieren percepción, cognición, planificación, aprendizaje, comunicación o acción física similares a las humanas; y tecnologías que pueden aprender y actuar de forma autónoma, ya sea en forma de agentes de software o robots incorporados.</i>

Fuente: Elaboración propia, con base en las definiciones expuestas previamente.

Estas definiciones posicionan los ingredientes que se deben estudiar para poder entender el impacto de los sistemas inteligentes en la vida cotidiana de una persona, comunidad o sociedad. Destaca la propuesta de Andrew W. Moore, Martial Hebert, Shane Shaneman, este último, como director de Investigación gubernamental estratégica - DoD / IC Carnegie Mellon University presenta lo que gráficamente es la pila de la IA:

⁴¹ Cfr. Comunicación de la Comisión Europea sobre la IA al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia Artificial para Europa, Bruselas, 25.4.2018 COM(2018) 237 final. Posteriormente evoluciona en la segunda definición, que podemos decir, incluye el *Machine Learning* y el *Deep Learning*.



Fuente: <https://ndiastorage.blob.core.usgovcloudapi.net/ndia/2019/solic/Shaneman.pdf>

Estas definiciones nos dan un panorama de la evolución conceptual que ha tenido la IA y nos muestran diversos ingredientes que se deben estudiar para poder comprender el impacto actual de los sistemas inteligentes (en la vida de la empresa, las personas, comunidades y, en conjunto, la sociedad).

A continuación, desarrollaremos la explicación de los dos tipos de IA que se estudian en nuestra investigación. El primero es la Inteligencia Artificial basada en Sistemas Expertos Decisionales (**SED**); y, el segundo es la Inteligencia Artificial basada en Machine Learning-Deep Learning (**ML-DL**).

1) Inteligencia Artificial basada en Sistemas Expertos Decisionales (SED)

Este tipo de IA puede ubicarse en lo que se conoce como Inteligencia Artificial “Débil o Estrecha”, aplicada a una tarea o un dominio específico; no puede aprender nuevas tareas o dominios, ni tomar decisiones por sí sola ni decisiones sin programación o sin datos de entrenamiento. Requiere de un programador y/o un experto en la cuestión a resolver que participan en la construcción de sus algoritmos.⁴² Tanto este tipo de IA como el que estudiaremos en el inciso B) de este apartado, se construyen con base en el “algoritmo”. En su definición tradicional, un algoritmo es un “conjunto ordenado y finito de operaciones que permite hallar la solución de un problema. Método y notación en las distintas formas del cálculo. Proviene, quizás, del latín tardío **algobarismus* y este abrev. del ár. clás. *ḥisābu lġubār* 'cálculo mediante cifras arábigas'.”⁴³ “Se piensa que el primer algoritmo no trivial es el algoritmo Euclídeo para el cálculo del máximo común divisor”.

Se dice, también, que el matemático persa Muhammad ibn Musa al-Khwarizmi (alrededor del año 820) introduce al mundo occidental el álgebra, los números arábigos y el “algoritmo”.⁴⁴

George Boole (1815-1864), considerado uno de los padres de la computación y de la lógica proposicional, “...presentó algoritmos para llevar a cabo deducciones lógicas y hacia el final del siglo XIX se llevaron a cabo numerosos esfuerzos para formalizar el razonamiento matemático general con la lógica deductiva.”⁴⁵

⁴² Existe también la llamada “Super AI o AI consciente” es IA con conciencia a nivel humano, lo que requeriría que sea consciente de sí mismo. Algunos autores comentan que, debido a que aún no podemos definir adecuadamente qué es la conciencia, es poco probable que podamos crear una IA consciente en el futuro cercano. Cfr. Russell y Norvig, REVISAR EN DÓNDE ESTÁ LA PRIMERA CITA DE ESTE LIBRO *op. cit.*, p. 26.

⁴³ Cfr. Voz: “Algoritmo”, *Diccionario de la lengua española*, edición del Tricentenario, 2020, en <https://dle.rae.es/algoritmo> [20 de julio de 2022]

⁴⁴ *Idem.*

⁴⁵ *Idem.* El álgebra de Boole está formada por un conjunto de variables con las que funcionan los sistemas de cómputo y eléctricos: 0, 1, abierto o cerrado, encendido o apagado, true - false.

Una definición de algoritmo vinculada con la Inteligencia Artificial es la de Christopher Steiner: “...una lista de instrucciones que lleva al usuario a una respuesta o salida en particular en función de la información disponible”.⁴⁶

El orden y la limitación (inicio y fin) son dos elementos claves del algoritmo (de su creación y utilización). Es decir, hay algo finito (tiene un inicio y un fin), que recibe información de entrada y produce una salida (resultados); y, debe estar bien definido (sin ambigüedades). La siguiente tabla sintetiza el panorama de definiciones que seleccionamos para efectos de acotar el concepto “algoritmo” a los fines de nuestra investigación:

Definiciones de algoritmo

Fuente	Definición
<i>Automate This: How Algorithms Came to Rule Our World</i>	The classical definition of an algorithm says the device is a list of instructions that leads its user to a particular answer or output based on the information at hand.
<i>Unboxing Artificial Intelligence: 10 steps to protect Human Rights. Recommendation</i> ⁴⁷	A finite suite of formal rules/commands, usually in the form of a mathematical logic, that allows for a result to be obtained from input elements.
<i>Algorithmic Accountability Policy Toolkit</i> ⁴⁸	Is generally regarded as the mathematical logic behind any type of system that performs tasks or make decisions... Algorithms do not have to be based in software on computers.
<i>Algorithmic Accountability: A primer</i> ⁴⁹	An algorithm is a set of instructions for how a computer should accomplish a particular task.

Fuente: Elaboración propia

La mayoría de los algoritmos son de dominio público y están en sitios abiertos para su descarga. Las empresas los han integrado como parte de sus sistemas informáticos, gracias a esos algoritmos la vida del hombre se ha vuelto más sencilla, ágil, dinámica. Las empresas han entendido que los algoritmos les dan ventajas competitivas. Ejemplos claros los vemos en Google, Uber, Akamai Technologies o Netflix. Google, por ejemplo, se basó en su algoritmo de PageRank para superar a otros competidores de motores de búsqueda. Uber, utiliza algoritmos de

⁴⁶ Steiner, Christopher, *Automate This: How Algorithms Came to Rule Our World*, Penguin Publishing Group. Edición de Kindle 2012, p. 5.

⁴⁷ Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights. Recommendation*, Council of Europe, Bruselas, 2019, p. 24.

⁴⁸ Commissioner for Human Rights, *Algorithmic Accountability Policy Toolkit*, October 2018, p.2.

⁴⁹ Caplan Robyn, Donovan, Joan, Hanson, Lauren and Matthews, Jeanna, *Algorithmic Accountability: A Primer*, Data & Society, Washington, p. 4.

enrutamiento para calcular la duración de los viajes y la tarifa que debe cobrar a los usuarios tomando en cuenta el tráfico del lugar. Akamai Technologies, un producto del MIT, se ha convertido en la plataforma más importante del mundo para distribuir contenido a través de Internet. Sus algoritmos dirigen las solicitudes de contenido de los usuarios a los servidores óptimos, lo que garantiza que la entrega sea rápida y por las rutas más eficientes. Netflix cuenta con un equipo de desarrolladores que trabajan para mejorar la descarga de videos en países o zonas donde el internet es deficiente, lo que le permite entregar contenidos de la misma calidad utilizando la mitad del ancho de banda.⁵⁰ A continuación, haremos un recorrido histórico para conocer el panorama general de los Sistemas Expertos Decisionales.

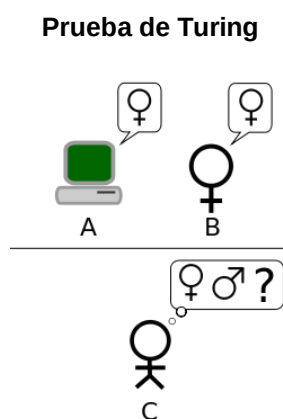
A) Evolución y conceptos cardinales

Fue Alan Turing (1912-1954), matemático inglés, quien publica en 1950 el artículo “Computing Machinery and Intelligence”, con el que pretende diferenciar entre entidades inteligentes y seres humanos. Turing sometió a la comunidad científica la pregunta: ¿Pueden pensar las máquinas? y abrió el debate con el llamado “juego de la imitación”⁵¹, al crear una prueba para determinar si una máquina podía

⁵⁰ Cfr. Neil C. Thompson, et al., “Building the Algorithm Commons: Who Discovered the Algorithms that Underpin Computing in the Modern Enterprise?”, *Global Strategy Journal*, no. 11, 2021, pp. 11-33. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/gsj.1393> [20 de julio de 2022]

⁵¹ Turing inicia su texto diciendo: 1. The Imitation Game. I PROPOSE to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. The new form of the problem can be described in terms of a game which we call the ‘imitation game’. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either ‘X is A and Y is B’ or ‘X is B and Y is A’. The interrogator is allowed to put questions to A and B thus: C: Will X please tell me the length of his or her hair? Now suppose X is actually A, then A must answer. It is A’s object in the game to try and cause C to make the wrong identification. His answer might therefore be ‘My hair is shingled, and the longest strands are about nine inches long.’ In order that tones of voice may not help the interrogator, the answers should be written, or better still, typewritten. The ideal arrangement is to have a

pensar. La máquina superaba la prueba de Turing si el evaluador (humano) no lograba distinguir entre las respuestas dadas por una persona y las respuestas dadas por una máquina. El evaluador se encontraba en una habitación distinta a la de la máquina y a la de la otra persona y sus respuestas le llegarían mediante otra máquina. Turing pensaba que una máquina se podía considerar inteligente si superaba la prueba, es decir, si el humano no podía distinguir el origen de las respuestas (humano o máquina).^{52[08]}



Fuente: https://en.wikipedia.org/wiki/File:Turing_Test_Version_1.svg

En 1956, Allen Newell, John Cliff Shaw y Herbert A. Simon, en su texto *Empirical Explorations of The Logic Theory Machine: A Case Study in Heuristic*, crean el programa de razonamiento “*Logic Theory Machine (LT)*”, capaz de pensar

teleprinter communicating between the two rooms. Alternatively, the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as ‘I am the woman, don’t listen to him!’ to her answers, but it will avail nothing as the man can make similar remarks. We now ask the question, ‘What will happen when a machine takes the part of A in this game?’ Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, ‘Can machines think?’ Cfr. Alan Mathison Turing, “Computing Machinery and Intelligence”, *Mind*, volumen LIX, número 236, octubre 1950, pp. 433-460. <https://doi.org/10.1093/mind/LIX.236.433> [15 de julio de 2022]

⁵² El cerebro, para algunos pensadores de esa época, podía ser visto como una máquina (ordenador o computadora) que a su vez controla a un cuerpo (robot). En palabras de John Searle, el fenómeno mental y su relación con el cuerpo tiene 4 rasgos característicos: Conciencia, Intencionalidad, Subjetividad y Causación. John Searle debate la prueba de Turing en: “Pueden los computadores pensar?” del libro *Mentes, Cerebros y Ciencia*, Ediciones Cátedra, 1985, Madrid, pp. 33-48. Consúltese también a Jack Copeland, *Inteligencia artificial*, Alianza Universidad, Madrid, 1996, pp. 69 a 96.

de manera no numérica. Los autores sistematizaron y codificaron la lógica simbólica. Este es el primer programa de inteligencia artificial creado para que las máquinas funcionaran con base en el razonamiento lógico. En dicho documento se puede encontrar el primer algoritmo que resolvía problemas imitando a los seres humanos (modelo simbólico computacional cognitivo):⁵³

En ese momento se utilizaban tarjetas de papel con perforaciones en una máquina física real cuya función era probar teoremas en lógica simbólica elemental. El cálculo se hacía con expresiones construidas a partir de combinaciones de símbolos básicos. Cinco de estas expresiones se tomaban como axiomas y existían reglas de inferencia para generar nuevos teoremas a partir de los axiomas y de otros teoremas. Las variables del sistema se interpretan como oraciones y los axiomas y reglas de inferencia como formalizaciones de operaciones lógicas. Se utilizan las variables p, q, r, A, B, C , y se hacían combinaciones con expresiones mediante conectivos. Por ejemplo, dada cualquier variable p , se puede construir la expresión “no- p ”. Dadas dos variables p y q , se forma la expresión “ p o q ”.

Características del lenguaje utilizado en *Logic Theory Machine (LT)*

Estructura	Características
Axiomas	$(p \text{ o } p) \text{ implica } p$ $p \text{ implica } (q \text{ o } p)$ $(p \text{ o } q) \text{ implica } (q \text{ o } p)$ $[p \text{ o } (q \text{ o } r)] \text{ implica } [q \text{ o } (p \text{ o } r)]$ $(p \text{ implica } q) \text{ implica } [(r \text{ o } p) \text{ implica } (r \text{ o } q)]$
Reglas de inferencia	Dados algunos teoremas verdaderos, se pueden derivar nuevos teoremas por medio de tres reglas de inferencia: - Sustitución: $(p \text{ o } q) \text{ implica } [q \text{ o } (p \text{ o } q)]$ - Reemplazo: no-p o $(q \text{ o } p)$ - Desprendimiento: $(p \text{ o } p) \text{ implica } p$, y $[(p \text{ o } p) \text{ implica } p] \text{ implica } (p \text{ implica } p)$, se obtiene un nuevo teorema: $p \text{ implica } p$.

Fuente: El autor con base en *Logic Theory Machine (LT)*

⁵³ Newell, Allen, Shaw John Cliff y Simon Herbert A., *Empirical Explorations of The Logic Theory Machine: A Case Study in Heuristic* en. E.A. Feigenbaum y J Feldman. Computers and Thought, Mc Graw-Hill, New York, 1963, pp. 109-133. Consúltese en: <http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=33559> [20 de julio de 2022]

Diagrama de flujo del LT

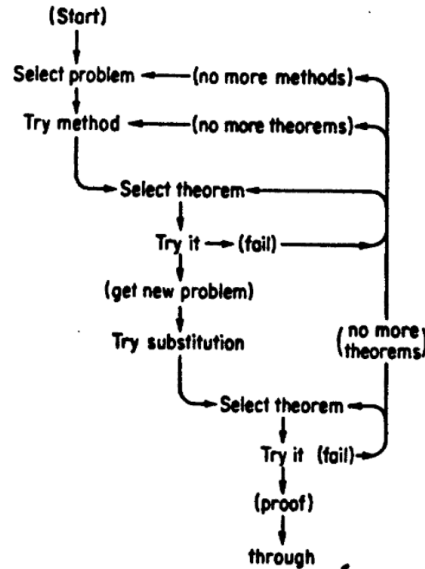


Figure 3. General flow diagram of LT.

Fuente: *Logic Theory Machine (LT)*

En el artículo se comprueba el rendimiento de un sistema de procesamiento de información que es capaz de encontrar pruebas de teoremas en lógica simbólica elemental, utilizado esos datos para analizar e ilustrar la diferencia entre los procesos algorítmicos sistemáticos, así como los procesos heurísticos de resolución de problemas. Demostrando cómo la *heurística*⁵⁴ puede servir al programa para poder resolver problemas en un tiempo de cálculo razonable, que, de otra manera, llevaría una mayor cantidad de tiempo.⁵⁵

No cabe duda de que el debate sobre el cerebro humano llevó a los científicos de esa época a retomar conceptos básicos de la lógica de Aristóteles (384 a. C – 322 a.C.), quien intentó codificar la “manera correcta de pensar” al

⁵⁴ La heurística, según el Diccionario de la Lengua Española es la “Técnica de la indagación y el descubrimiento” en <https://dle.rae.es/heur%C3%ADstico#KHdGTfC> Gabriel Zaid nos comenta que Herbert A. Simon propuso en 1957 el desarrollo de programas de computación heurística para que las máquinas tomen decisiones, a partir de criterios mecánicos (*The new science of management decision*) en su artículo: Heurística, *Letras Libres*, 7 de marzo de 2013, México. Revista Digital: <https://www.letraslibres.com/mexico-espana/heuristica-0> [20 de julio de 2022]

⁵⁵ Newell, Allen, Shaw John Cliff y Simon Herbert A., *op. cit.*, p. 133. Consúltase en: <http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=33559> [20 de julio de 2022]

construir los llamados “silogismos” o esquemas de estructura argumentativa constituida por una premisa mayor, una premisa menor y la conclusión. Este tipo de razonamiento dio las bases para crear la “lógica”, es decir, “...el estudio de los métodos y principios que se usan para distinguir el razonamiento bueno (correcto) del malo (incorrecto).⁵⁶

Esta forma de ver el pensamiento humano y los intentos por que las máquinas lo replicaran llevó a una división de la IA en lo que se conoce como modelos conexionistas y modelos simbólicos. Los primeros, buscan desarrollar procesos inteligentes copiando la estructura del cerebro (hace una combinación de elementos físicos y virtuales); los segundos, pretenden construir sus aportaciones a partir del estudio de las leyes del pensamiento y el lenguaje de símbolos para resolver problemas de manera similar al ser humano.⁵⁷ Esta división también sirve para diferenciar la IA del inciso **A)** de este apartado (con raíces muy marcadas en la lógica) y la IA del inciso **B)** basada en la biología, en las neuronas y en algunos otros conceptos del funcionamiento de los átomos y las reacciones químicas.

Los científicos que han estudiado estos temas se preguntan cuál debe ser el mejor el lenguaje de comunicación con las máquinas o cómo lograr que las máquinas hablen el lenguaje de los seres humanos. Los lenguajes de programación son las herramientas con las que el ser humano instruye a las máquinas lo que deben hacer. Sin estos, no hubiera sido posible el funcionamiento y evolución de la tecnología que hoy tenemos. El lenguaje de programación es el medio de comunicación por el cual las máquinas reciben instrucciones con el fin de lograr determinado objetivo. Esto se produce con la escritura de un código, que el

⁵⁶ Irving M. Copi y Carl Cohen, *Introducción a la lógica*, Limusa, México, 2013, p. 17 y ss.

⁵⁷ Hubo una prevalencia de modelos simbólicos durante todo el siglo XX. Sin embargo, actualmente estamos viendo que los modelos conexionistas están brindando mayores resultados para determinadas actividades como vehículos autónomos, reconocimiento de imágenes, predicciones, entre otros. Estamos dejando la precisión de la lógica por la percepción de los modelos de redes neuronales profundas.

dispositivo interpreta. Existe una tendencia a clasificar los lenguajes de programación⁵⁸ en:

- *Lenguaje Máquina*: Basado en el lenguaje binario de 0 y 1.
- *Lenguaje de Bajo nivel*: Ensamblador.
- *Lenguaje de Alto nivel*: Pascal, FORTRAN, Visual Basic, C, Ada, Modula-2, C++, Java, Delphi, C#, Python.
- *Lenguaje para Diseño Web*: SMGL, HTML, XML, PHP.

Código binario o código máquina

Letra	Decimal	Binario	Letra	Decimal	Binario	Letra	Decimal	Binario
—	0	00000	I	10	01010	Q	20	10100
A	1	00001	J	11	01011	R	21	10101
B	2	00010	K	12	01100	S	22	10110
C	3	00011	L	13	01101	T	23	10111
Ch	4	00100	LI	14	01110	U	24	11000
D	5	00101	M	15	01111	V	25	11001
E	6	00110	N	16	10000	W	26	11010
F	7	00111	Ñ	17	10001	X	27	11011
G	8	01000	O	18	10010	Y	28	11100
H	9	01001	P	19	10011	Z	29	11101

Fuente: <https://sites.google.com/site/valeriexmoon747/home/codigo-binario> [julio de 2022]

⁵⁸ Joyanes Aguilar, Luis. *Fundamentos de programación. Algoritmos, estructura de datos y objetos*, McGrawHill, Cuarta Edición, Madrid, 2004, p. 39-43.

Ejecución y fases de un programa

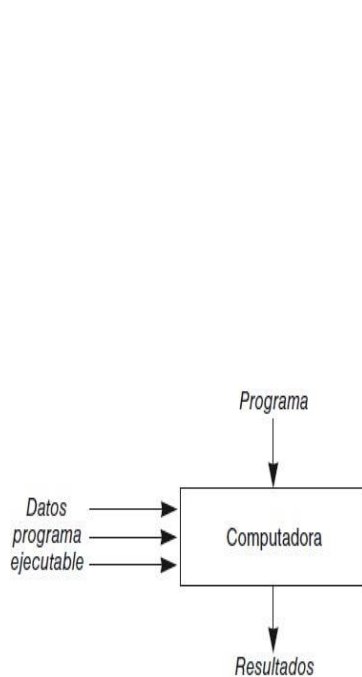


Figura 1.19. Ejecución de un programa.

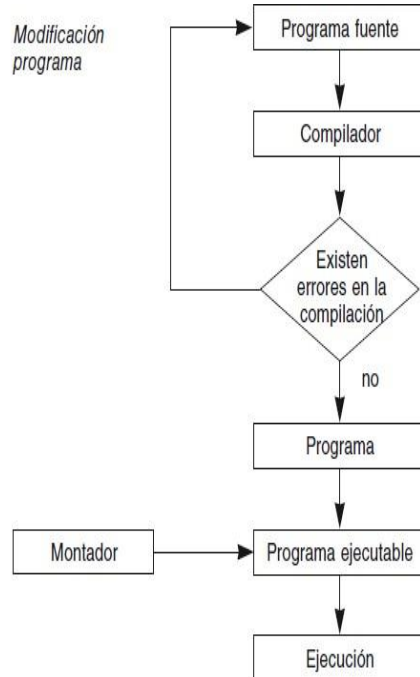


Figura 1.20. Fases de ejecución de un programa.

Fuente: Joyanes Aguilar, Luis. *Fundamentos de programación. algoritmos, estructura de datos y objetos*, McGraw Hill, Cuarta Edición, Madrid, 2004, p. 39.

Para entender la importancia de los lenguajes de programación, tengamos presente que las máquinas funcionan con energía eléctrica y que el “lenguaje máquina” conocido como “lenguaje binario” deriva de la energía que reciben y con el cual procesa las instrucciones basadas en “bits” (0 y 1); cero quiere decir que no pasa corriente (negativo-falso) y el uno que sí pasa corriente (positivo-verdadero). De tal forma que, si se presiona la tecla C en una máquina, su equivalente en sistema binario sería 01000 0011. El código internacional de equivalencias internas del sistema binario se conoce como “ASCII” (*American Standard Code for Interchange Information*).

En un nivel superior del lenguaje binario o lenguaje máquina está el lenguaje de bajo nivel, o lenguaje ensamblador, que utiliza cada procesador y las instrucciones que rescriben se hacen en códigos alfabéticos (nemotécnicas) usando

palabras en inglés para hacer el vínculo con el lenguaje máquina⁵⁹. Adicional a estos lenguajes, existe el lenguaje de alto nivel en el que se utilizan palabras para construir las instrucciones. El primer lenguaje de alto nivel que se creó fue FORTRAN, el segundo LISP y el tercero COBOL. Y, los tres lenguajes de programación “más utilizados han sido FORTRAN, Pascal y C, que representan el arquetipo de la programación procedimental, también relacionada estrechamente y —normalmente— conocida como programación estructurada. La programación con soporte en C y Pascal proporciona el paradigma procedimental tradicional con un énfasis en funciones, plantillas de funciones y algoritmos genéricos.”⁶⁰

La siguiente tabla contiene un listado de algunos lenguajes de programación:

⁵⁹ Luis Joyanes Aguilar hace una precisión en el uso de los conceptos: “Los lenguajes de programación se utilizan para escribir programas. Los programas de las computadoras modernas constan de secuencias de instrucciones que se codifican como secuencias de dígitos numéricos que podrán entender dichas computadoras. El sistema de codificación se conoce como lenguaje máquina que es el lenguaje nativo de una computadora. Desgraciadamente la escritura de programas en lenguaje máquina es una tarea tediosa y difícil ya que sus instrucciones son secuencias de 0 y 1 (*patrones de bit*, tales como 11110000, 01110011..) que son muy difíciles de recordar y manipular por las personas. En consecuencia, se necesitan lenguajes de programación “amigables con el programador” que permitan escribir los programas para poder “charlar” con facilidad. con las computadoras. Sin embargo, las computadoras sólo entienden las instrucciones en lenguaje máquina, por lo que será preciso traducir los programas resultantes a lenguajes de máquina antes de que puedan ser ejecutadas por ellas. Cada lenguaje de programación tiene un conjunto o “juego” de instrucciones (acciones u operaciones que debe realizar la máquina) que la computadora podrá entender directamente en su código máquina o bien se traducirán a dicho código máquina. Las instrucciones básicas y comunes en casi todos los lenguajes de programación son: • *Instrucciones de entrada/salida*. Instrucciones de transferencia de información entre dispositivos periféricos y la memoria central, tales como “leer de..” o bien “escribir en..”. • *Instrucciones de cálculo*. Instrucciones para que la computadora pueda realizar operaciones aritméticas. • *Instrucciones de control*. Instrucciones que modifican la secuencia de la ejecución del programa. Cfr. *Fundamentos de programación. Algoritmos, estructura de datos y objetos*. McGrawHill, Cuarta Edición, Madrid, 2004, p. 36-37.

⁶⁰ Ibidem., p. XVIII.

Lenguajes de programación

Lenguaje	Características
FORTRAN	Es considerado el primer lenguaje de programación de “alto nivel”. FORTRAN (Formula Translating System, 1954) fue usado en la primera máquina de cálculos complejos. Fue desarrollado por IBM para el equipo IBM 704, la cual se programaba recibiendo sus secuencias de instrucciones a través de tarjetas de papel, en las cuales las instrucciones iban perforadas, transfiriéndose de un registro a otro por medio de señales eléctricas.
LISP	John McCarthy, uno de los padres de la IA, crea el lenguaje “LISP” (<i>List-Processor</i> , 1958); el cual servirá para programar a los sistemas expertos decisionales. Se basa en modelos algebraicos, cada función o actividad viene en forma de lista.
COBOL	COBOL (Common Business-Oriented Language, 1959), que fue desarrollado por la U.S. Navy de Estados Unidos, para aplicaciones de gestión o administración.
C	El lenguaje C fue creado en Bell Laboratories (1971) para su uso científico y de gestión. Es estructurado, se pueden escribir bucles que tienen condiciones de entrada y salida claras y se pueden escribir funciones cuyos argumentos se verifican siempre para su completa exactitud. Cuenta con biblioteca. Sin embargo, carece de orientación a objetos.
C++	Creado en 1983, es un tipo de programación orientada a objetos. C++ es uno de los lenguajes más utilizados en el mundo, y es por eso que tiene un lugar muy merecido en la lista de los lenguajes de programación más usados. Google Chrome, Mozilla Firefox, Winamp, y todas las aplicaciones de Adobe Software fueron desarrolladas usando C++. ⁶¹
JAVA	Java fue inventado en 1995 por James Gosling, de la compañía SUN (adquirida el 27 de enero de 2010 por la compañía Oracle). Es una evolución de la sintaxis de los lenguajes de programación C y C++. Las aplicaciones de Java son compiladas como bytecode (clase Java), que puede ejecutarse en cualquier máquina virtual Java (JVM) sin importar la arquitectura de la computadora. Fue pensado para lograr lo siguiente: Ser un modelo para la programación orientada a objetos. Permitir la ejecución de un mismo programa en múltiples sistemas operativos. Soportar el trabajo en red. Ejecutar código remoto de forma segura. Fácil de usar.
Python	Su lanzamiento se dio en 1991 bajo la creación de Guido van Rossum, es un lenguaje de código abierto. Frente a otros lenguajes como Java o C, tiene la ventaja que es un lenguaje interpretado y no necesita compilación previa, facilitando el ciclo de programación y pruebas. Es un lenguaje de propósito general que puede utilizarse para el desarrollo de cualquier proyecto; es un lenguaje multiplataforma, por tanto, lo que se cree puede funcionar en cualquier dispositivo y en cualquier sistema operativo. Cuenta con un “interprete interactivo” que nos permite ejecutar instrucciones directamente sin necesidad de escribirlas dentro del programa. Puede trabajar con secuencias de datos de forma natural y directa. En otros lenguajes hay que recurrir a APIs para manejo de estructuras de datos. Se integra por listas, tuplas, conjuntos y diccionarios. Se pueden crear cualquier tipo de clases para cualquier tipo de objeto. Los datos, funciones y definiciones se organizan en módulos. Hay la opción de separar los módulos por directorios para diferenciar cosas.

Fuente: El autor con base en: Joyanes Aguilar, Luis. *Fundamentos de programación. Algoritmos, estructura de datos y objetos*, Mc Graw Hill, Cuarta Edición, Madrid, 2004 p. 42. Consultar también: <https://www.python.org/> [15 julio de 2022]

⁶¹ Consultar: <https://classpert.com/blog/lenguajes-de-programacion-mas-usados?intl=true> [20 de julio de 2022]

El lenguaje de programación de una computadora, conjuntamente con los programas o software⁶² (cada vez más potentes y fáciles de utilizar), han permitido el desarrollo de sistemas como Watson. La programación orientada a objetos (como la de C++ y Python) están siendo utilizada para diversas aplicaciones relacionadas con ML para resolver problemas complejos. A diferencia de la programación procedimental, es decir, la que tiene su fortaleza en los algoritmos tradicionales (programados por el ser humano), la programación orientada a objetos pone más énfasis en los datos, intentando ajustar el lenguaje al problema. El uso de los datos, bajo formatos específicos, que se construyen tomando en cuenta la esencia del problema debe combinar tanto los datos como las funciones en un mismo módulo. Tal módulo se conoce como “funciones miembro o métodos” y son el medio para acceder a los datos y las propiedades ligadas a la orientación de objetos. Por lo general, se relacionan con: 1. Abstracción (tipos de datos y clases); 2. Encapsulado de datos; 3. Ocultación de datos; 4. Herencia; y 5. Poliformismo.⁶³ Luis Joyanes Aguilar comenta:

Los datos de un objeto se conocen también como *atributos* o *variables de instancia*. Si se desea leer datos de un objeto, se llama a una función miembro del objeto. Se accede a los datos y se devuelve un valor. No se puede acceder a los datos directamente. Los datos están ocultos, de modo que están protegidos de alteraciones accidentales. Los datos y las funciones se dice que están *encapsulados en una única entidad*. El *encapsulamiento de datos* y la *ocultación* de los datos son términos clave en la descripción de lenguajes orientados a objetos. Si se desea modificar los datos de un objeto, se conoce exactamente cuáles son las funciones que interactúan con miembros del objeto. Ninguna otra función puede acceder a los datos. Esto simplifica la escritura, depuración y mantenimiento del programa. Un programa C++ se compone normalmente de un número de objetos que se comunican unos con otros mediante la llamada a otras funciones miembro.⁶⁴

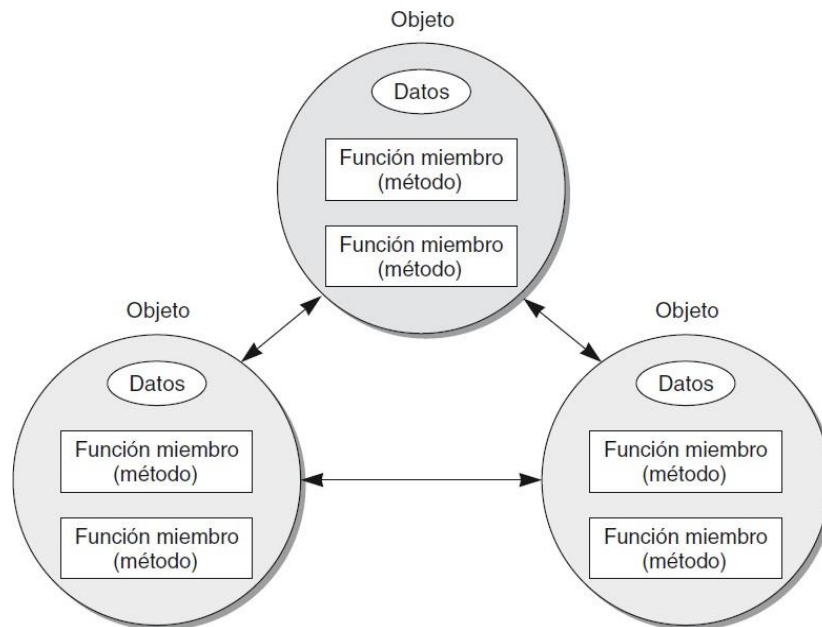
⁶² Luis Joyanes Aguilar explica lo siguiente “Un programa de software es un conjunto de sentencias o instrucciones a la computadora. El proceso de escritura o codificación de un programa se denomina programación y las personas que se especializan en esta actividad se denominan programadores. Existen dos tipos importantes de software: software del sistema y software de aplicaciones. Cada tipo realiza una función diferente. El software del sistema es un conjunto generalizado de programas que gestiona los recursos de la computadora, tal como el procesador central, enlaces de comunicaciones y dispositivos periféricos. Los programadores que escriben software del sistema se llaman programadores de sistemas. El software de aplicaciones es el conjunto de programas escritos por empresas o usuarios individuales o en equipo y que instruyen a la computadora para que ejecute una tarea específica. Los programadores que escriben software de aplicaciones se llaman programadores de aplicaciones.” Ibidem., p. 32.

⁶³ Para un desarrollo detallado de cada propiedad consúltese a Luis Joyanes Aguilar, *op. cit.*, p. 57 y ss.

⁶⁴ Ibidem., p. 57.

La siguiente imagen muestra la organización típica de un programa C++ orientado a objetos:

Programación orientada a objetos



Fuente: Luis Joyanes Aguilar, *Fundamentos de programación. Algoritmos, estructura de datos y objetos* p. 57.

En este sentido, hacer “programación” comprende el análisis de una problemática determinada, y a la vez, diseñar una solución para la misma, desarrollarla en el programa para atenderla; realizar pruebas antes de ponerla en fase de producción y realizar correcciones.

La experiencia nos dice que un requisito que debe llenar quien desee construir un sistema o el desarrollo de alguna aplicación para el uso de inteligencia artificial está en la detección con claridad del problema y la creación de una solución que sea programable. Un programa bien estructurado utiliza “bloques de información” con instrucciones precisas para su uso. Existen tres tipos de bloques para programar cualquier aplicación:

- El bloque secuencial, es decir, una sentencia lleva a la otra (es el más simple).

- El bloque repetitivo (bucle o iteración), efectúa la ejecución iterada de una sección del programa.
- El bloque de selección sirve para realizar secuencias diferentes de instrucciones según la condición.

La utilización de todo lenguaje de programación conlleva a cuidar los errores humanos. Los errores más comunes en la programación son:

- Errores de concepción: cuando el programador cree conocer bien el objetivo, sin embargo, no es así. Finaliza el programa con algo que no resuelve el problema.
- Errores lógicos: cuando los algoritmos no están bien contruidos y presentan fallas en su orden o en su inicio y fin.
- Errores de procedimiento: cuando se cree que el algoritmo está bien, se hace la prueba de escritorio para verificarlo y se cree que funciona bien.
- Errores de sintaxis: cuando realizamos la programación podemos tener errores en la transcripción
- Errores semánticos: cuando la sintaxis del código está bien articulada, pero el significado no es el que se buscaba.⁶⁵

El debate entre el funcionamiento del cerebro humano, la inteligencia artificial y el lenguaje fueron incorporando a diversas disciplinas, como por ejemplo de la Filosofía (dualismo, materialismo, empirismo, inducción, positivismo, lógica, teoría de la confirmación); las matemáticas (algoritmos, teorema de incompletitud, probabilidad); la economía (teoría de la decisión, teoría de juegos, satisfacción y utilidad); neurociencia (cómo procesa el cerebro la información; las neuronas; inteligencia, conciencia); la psicología (conductismo, psicología cognitiva); la Ingeniería computacional; la cibernética (teoría de control); lingüística (lenguaje y pensamiento; representación del pensamiento); y con ellas, la creación de seis actividades científicas que se relacionan con las funciones que una máquina puede realizar:

⁶⁵ Trejos Buriticá, Omar, Lógica de programación, Medellín, Ediciones de la U, Bogotá, 2017, pp. 133 a 135.

- Procesar el lenguaje natural (sistemas expertos).
- Representar conocimiento (sistemas expertos).
- Generar razonamiento automático (sistemas expertos).
- Aprender automáticamente para adaptarse a nuevas circunstancias y patrones (ML/DL).
- Desarrollar visión computacional (ML/DL).
- Manipular objetos (robótica).

Una vez que un programa informático se escribe en un idioma que los humanos comprenden, como C ++, debe traducirse a un lenguaje de nivel inferior que las máquinas comprendan; este proceso de traducción se conoce como “compilación”. Alfred Aho y Jeffrey Ullman obtuvieron el Premio Turing 2020, considerado como el Nobel de la Informática por ayudar a describir la tecnología de los compiladores y ampliar la traducción de programas de alto nivel en códigos máquina de bajo nivel. Todo lo anterior nos lleva al uso de “Sistemas Expertos Decisionales”, su construcción, funcionamiento, tipos y su relación con el procesamiento del lenguaje natural, la representación de conocimiento y la generación el razonamiento automático por un experto humano. En la siguiente sección explicaremos a detalle todos estos temas.⁶⁶

B) Sistemas Expertos Decisionales (SED)

Los SED son una rama de la Inteligencia Artificial tradicional que surgieron en la década de 1960. Un SED es capaz de “decidir/recomendar” con base en reglas, conocimiento, casos y/o la experiencia de un humano determinada solución.⁶⁷ El

⁶⁶ Cfr. Myers, Andrew, “Stanford’s Jeffrey Ullman receives ACM Turing Award”, *Stanford News*, del 31 de marzo de 2021, en <https://news.stanford.edu/2021/03/31/jeffrey-ullman-receives-acm-turing-award/> Para conocer más sobre la opinión de expertos en lenguajes de programación y su evolución consúltese: Herranz, Arantxa, “20 años después quiero volver a programar: qué grandes cambios ha habido y por dónde empezar, XATAKA, 22 de junio de 2020. Consúltese en: <https://www.xataka.com/otros/20-anos-despues-quiero-volver-a-programar-que-grandes-cambios-ha-habido-donde-empezar> [20 de julio de 2022]

⁶⁷ Cfr. Edward A. Feigenbaum y Philip Klahr, “Expert Systems”, *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., Londres, 2013, pp. 684–689. Consúltese en: <https://dl.acm.org/doi/10.5555/1074100.1074389> [20 de julio de 2022]

objeto de los SED está en el aprovechamiento del conocimiento y experiencia de un humano para transferirlo a la computadora. La computadora almacena el conocimiento, el usuario humano le pide consejos específicos y la computadora hace inferencias hasta llegar a una conclusión específica.

Para ello, se pensó que era necesario buscar a expertos humanos especializados en temas determinados para que respondieran a preguntas sobre casos recurrentes. Uno de los primeros SED es el sistema basado en reglas. Las reglas almacenadas en el sistema tienen forma de Antecedente = Consecuente (IF/THEN)⁶⁸:

R001 SI tiene dificultad de respirar y pitidos en los bronquios ENTONCES el paciente sufre de asma

R002 SI tiene dolor muscular y tos y fiebre, ENTONCES el paciente tiene gripe.

La segunda categoría de los SED encontramos a los “Sistema Basado en Conocimiento”. El programa DENDRAL (*Dendritic Algorithm*), creado en 1965 por Ed Feigenbaum, Bruce Buchanan y Joshua Lederberg, es considerado el primer sistema de conocimiento formado por una base de información hecha por expertos y reglas para su aplicación. Feigenbaum junto con otros investigadores de la Universidad de Stanford, al entender la importancia de la construcción de sistemas de conocimiento comenzaron el proyecto de programación heurística.

La idea que tenían en mente sus creadores era determinar la nueva metodología que se utilizaría para los “sistemas expertos”. Feigenbaum, Buchanan y Shortliffe diseñaron el programa MYCIN para el diagnóstico de enfermedades sanguíneas⁶⁹. MYCIN generaba diagnósticos tan precisos como un médico experto y mucho mejores que un estudiante de medicina. MYCIN funcionaba bajo tres premisas:

⁶⁸ García Serrano, Alberto, *Inteligencia artificial: Fundamentos, práctica y aplicaciones*, Alfaomega, México, 2016, p. 172 y ss.

⁶⁹ Rusell y Norvig, *op. cit.*, p. 27. También consúltese a: Edward A. Feigenbaum y Philip Klahr, “Expert systems”, *Encyclopedia of Computer Science*. John Wiley and Sons Ltd., Londres, 2003, pp. 684–689.

1) Un sistema de consulta, que utiliza la información proporcionada por el médico, junto con su propio conocimiento basado en reglas, para elegir un régimen terapéutico apropiado para un paciente con una infección bacteriana;

2) Un Sistema de Explicación, que comprende preguntas sencillas y las responde para justificar sus decisiones o instruir al usuario;

3) Un sistema de adquisición de reglas, que acepta nuevas reglas de expertos y las codifica para su uso en futuras sesiones de consulta.⁷⁰

En 1978 se crea el primer sistema experto comercial. Se denominó “R1” y fue creado por *Digital Equipment Corporation* y la Universidad de Carnegie-Mellon. Fue utilizado para el control de pedidos de sistemas informáticos. Utilizaba el lenguaje de programación OPS5, mediante el cual, el sistema experto interactuaba con el departamento de ventas y configuraba los pedidos haciendo determinadas preguntas antes de imprimir la hoja de venta final. Se comprobó que le generaba a la compañía un ahorro considerable.

Durante la década de 1980, la proliferación de computadoras personales y la introducción de herramientas para la construcción de los SED (denominadas *expert system shells* o intérprete de órdenes o intérprete de comandos) hizo que su uso se multiplicara exponencialmente; y en la década de 1990, la complejidad y el volumen del conocimiento médico aumentaron, encontrándose aproximadamente un total de 233 SED entre 1992 y 1996.⁷¹

La siguiente tabla muestra una breve descripción de los más comunes *expert system shells* que existen actualmente, si son de código abierto, si adoptan el

⁷⁰ Edward H. Shortliffe, “A rule-based computer program for advising physicians regarding antimicrobial therapy selection” en *Proceedings of the 1974 annual ACM conference - Volume 2* (ACM '74), Association for Computing Machinery, New York, p. 739. DOI: <https://doi.org/10.1145/1408800.1408906> [20 de julio de 2022]

⁷¹ S. R. Soni, A. Khunteta, and M. Gupta, “A review on intelligent methods used in medicine and life science” en *Proceedings of the International Conference & Workshop on Emerging Trends in Technology* (ICWET '11). Association for Computing Machinery, New York, 2011, pp. 703–706. DOI: <https://doi.org/10.1145/1980022.1980173> [20 de julio de 2022]

estándar JSR-94⁷² y en qué lenguaje de programación están contruidos. Estos *shells* brindan la posibilidad de administrar e interpretar el conocimiento como reglas y hechos.

Table 1. Common expert systems / rule engines.

	active dev	open-source	jsr-94	language
JavaDON		✓		java
NxBRE	✓	✓		c#
JESS	✓		✓	java
CLIPS	✓	✓		c
JBoss Drools	✓	✓	✓	java
OpenRules	✓	✓	✓	java
ILog JRules	✓		✓	java
Pyke		✓		python
OpenExpert		✓		php
DTRules		✓		java
Infosapient				java
JEops				java
JLisa			✓	java
Mandarax		✓		java
OpenCyc	✓			java

Fuente: Stefan Nadschläger y Josef Küng, "Towards a pattern language for knowledge processing systems: expert systems" en *Proceedings of the VikingPloP 2017 Conference on Pattern Languages of Program (VikingPloP '17)*, Association for Computing Machinery, New York, 2017, article 1, 1–33. DOI: <https://doi.org/10.1145/3158491.3158506> [15 julio de 2022]

En los SED encontramos el primer intento para resolver problemas o tomar decisiones de manera especializada similar al razonamiento humano. Estos programas se pensaron para utilizar la *expertise* de una persona (en un área del conocimiento) para la toma de decisiones que tiene que realizar otra (basada en inferencias). Utilizan algunos algoritmos como base (que permiten la operación del antecedente y consecuente) para procesar la información y las reglas de la lógica o de alguna otra metodología. Para entender más sobre su funcionamiento, a continuación, se explican algunos modelos tradicionales:

1. **INTERNIST:** Se inició a principios de la década de 1970 y continúa hoy bajo el nombre de CADUCEUS. El objetivo de INTERNIST es realizar un

⁷² JSR-94 o *Java Specification Request* (JSR) 94, es un API del motor de reglas de Java para SED, contiene las mejores prácticas para la implementación de un motor de reglas y proporciona una definición de los componentes básicos, así como una funcionalidad para manejo de excepciones, registro de errores e implementación. Esta API no se puede aplicar a los sistemas de procesamiento del conocimiento en general.

diagnóstico de la mayoría de las enfermedades asociadas con el campo de la medicina interna. INTERNIST / CADUCEUS no sólo está destinado a diagnosticar cada enfermedad, sino que está diseñado para considerar todas las combinaciones posibles de enfermedades que podrían estar presentes en el paciente.

2. *Exsys Corvid*: Este sistema funciona desde 1983. El objetivo de los sistemas expertos en automatización del conocimiento como el *Exsys Corvid* es capturar, automatizar y difundir consejos y recomendaciones para la toma de decisiones de cualquier empresa. EXSYS Inc. fue el primero en llevar sistemas interactivos de asesoramiento a la Web (1996). En lugar de analizar solamente datos, los sistemas interactivos Exsys brindan beneficios mediante el uso de sitios web y equipos de cómputo independientes al proporcionar respuestas lógicas que se aplican específicamente a cada usuario. Los pasos para construir un sistema son muy similares a explicarle a una persona cómo se toma una decisión. Las herramientas integradas de CORVID ayudan a los desarrolladores a pensar y describir sus pasos para la toma de decisiones y la lógica de resolución de problemas. Los sistemas CORVID de Exsys se pueden entregar a través de la Web, que crea dinámicamente páginas HTML para la interfaz de usuario, o mediante Java Applet Runtime que se muestran en la máquina del cliente. Las plantillas CORVID Servlet Runtime permiten utilizar la flexibilidad total de HTML para combinar el diseño del sitio. Se pueden hacer varias preguntas en una sola pantalla. Proporcionar respuestas expertas interactivas, no solo información, es fundamental para ayudar a los prospectos, clientes y empleados a tomar decisiones rápidas, correctas y consistentes. Algunas de las áreas más comunes para la automatización del conocimiento que atiende son:
 - a) Atención al cliente: sitios de “autoservicio”.
 - b) Solución de problemas: diagnóstico de procesos y reparación de maquinaria.

- c) Cumplimiento: comprensión y conformidad con las reglamentaciones, la política de la empresa y las “mejores prácticas”.
 - d) Selección de productos: asesoramiento de ventas para recomendar los mejores productos para cada cliente.
 - e) Cuestionarios inteligentes: recopilación de datos de manera efectiva, en una forma única interactiva y fácil de usar.
 - f) Mantenimiento predictivo: detección de problemas antes de que se vuelvan irreparables o costosos.
 - g) Programación: evitar problemas de asignación de recursos.
 - h) Informes automatizados: análisis exclusivo de datos complejos y generación de informes.
 - i) Detección de fraude: compruebe los datos y las situaciones en busca de inconsistencias.⁷³
3. *NEOTALIGIC*: Es un Sistema Experto que está evolucionando hacia el ML y el DL. Pero que mantiene varios de sus productos vinculados al modelo de IA tradicional. En el área legal dan soporte a despachos grandes con soluciones basadas en sistemas expertos. Su cofundador Michael Mills ha comparado sus resultados con el sistema Watson en su sección Legal. Sin embargo, afirma que Watson Legal no ha podido ser aplicado de forma exitosa en ningún despacho. Otro competidor es ROSS Intelligence que utiliza a Watson para este sector de servicios y utilizar sus aportaciones en técnicas de procesamiento del lenguaje natural. Neota Logic aplica su plataforma de razonamiento híbrido, que combina sistemas expertos y otras técnicas como la PNL bajo demanda y el aprendizaje automático, para proporcionar respuestas específicas de hechos y contextos legales, de cumplimiento y políticas públicas.⁷⁴

⁷³ Para mayor información consúltase: <http://www.exsys.com/productmain.html> y <http://www.exsyssoftware.com/CaseStudySelector/casestudies.html> [20 de julio de 2022]

⁷⁴ Para conocer otras empresas de IA que están dando soporte al sector legal, se puede consultar el siguiente artículo: Caroline, Hill, “Artificial Intelligence in Law – The State of Play in 2015?”, *Legal Technology*, 3 de noviembre de 2015. Consúltase en: <https://legaltechnology.com/artificial-intelligence-in-law-the-state-of-play-in-2015/> Existen también otros sistemas expertos como: *COMPLIANCEHR* (<https://Compliancehr.com/>) y el Sistema Experto Open Net (<https://docassemble.org/> y <https://docassemble.org/docs/legal.html>) [20 de julio de 2022]

4. *IBM Pure Application System*: Estos sistemas expertos fueron absorbidos por Watson. Buscó integrar servidores de almacenamiento y redes para facilitar la gestión de las empresas (pequeñas, medianas y grandes), se entregaban patrones de experiencia de más de 150 sectores, para que el usuario los utilizara con tan solo descargar el producto. Fueron el primer escalón para la transición hacia el *cloud computing*, principalmente en España, Portugal, Grecia e Israel. IBM entendió que el uso de patrones en determinadas industrias simplificaba las actividades cotidianas y se podían automatizar, reduciendo costos y tiempo por todo el ciclo de vida de la aplicación. Se ofrecían patrones para aplicaciones, para plataformas y para infraestructura.⁷⁵

Primeros Sistemas Expertos Decisionales

Año	Tipo de sistema
1965	DENDRAL (Dendritic Algorithm) sobre estructuras químicas MACSYMA (análisis matemáticos) HEARSAY (lenguaje)
1972	MYCIN (diagnóstico de enfermedades de la sangre) TIERESIAS (herramienta para transformar conocimientos diversos) PROSPECTOR (exploración mineral y herramientas de identificación)
1973	AGE (generador de sistemas expertos)
1974	OPS5 (generador de sistemas expertos)
1975	CADUCEUS (herramienta para medicina interna)
1976	COURTRAN (operativa criminal, investigación criminal, operativa civil e investigación civil)
1978	R1 (configuración de pedidos de venta de equipos de cómputo)

Fuente: El autor. Los lenguajes de programación que utilizan estos sistemas son APL, PROLOG y LISP.

De manera general, podemos decir que un SED -BÁSICO- se integra por:

- a) Base del conocimiento (hecha por un experto): La constituye el conocimiento que aporta el experto humano; éste es el responsable de dotar del conocimiento y se relaciona con la especialidad del tema que se busca atender a través del sistema experto decisional; El experto suministra la información que se procesará por el programa. Ayuda a construir las reglas y verifica la relación lógica de la información (antecedente / consecuente).

⁷⁵ Pure Application System, para mayores datos consúltense: https://www.ibm.com/ibm/puresystems/uy/es/pf_pureapplication.html [20 de julio de 2022]

- b) Base de hechos: Es el archivo que guarda la memoria de la información (los hechos). Los sistemas expertos pueden trabajar con grandes volúmenes de información.
- c) Motor de inferencia: Sirve para llegar a conclusiones aplicando el conocimiento y los datos. Existen dos tipos: Encadenamiento hacia adelante (a partir de los hechos disponibles, infiere las conclusiones posibles). Encadenamiento hacia atrás (a partir de un objetivo intenta verificar los hechos que lo sostienen) El encadenamiento de reglas se utiliza cuando las premisas de ciertas reglas coinciden con las conclusiones de otras. Cuando se encadenan las reglas, los hechos pueden utilizarse para dar lugar a nuevos hechos. Se repite sucesivamente hasta llegar a más conclusiones. Los especialistas suministran al programador la base de conocimiento de forma ordenada y actualizada y bajo tres categorías: Reglas, Afirmaciones y preguntas; y, Hechos y sus relaciones. El mecanismo de inferencia revisa la base de datos del conocimiento para encontrar la respuesta a una pregunta desde la interfaz del usuario. Controla el diálogo con el usuario, determina cómo y cuándo se procesan las reglas; acciones y el orden en cómo intervendrá.
- d) Metodología: Existe una serie de pasos que se deben tener presentes para la construcción de un sistema experto. Se utilizan algunas preguntas guía, como, por ejemplo: ¿Cuál es la problemática que debemos resolver? Elaborar problemario (responder a preguntas / resolver problemas) ¿Quiénes son los mejores expertos al alcance? Ubicar candidatos para que desarrollen la solución al problema ¿Con qué hardware contamos? (lugar de almacenamiento, motor de inferencia, interfaz de usuario) ¿Qué software utilizaremos? Elección de la herramienta para su desarrollo o lenguaje de programación (C++, Java, Python, otros) ¿Tiene una ruta crítica para su desarrollo? Prueba y corrección ¿Sabe el costo de su uso? Mantenimiento y actualización.⁷⁶

⁷⁶ José Luis, Hernández Sánchez, *Sistemas Expertos Decisionales*, Diplomado Ciberjusticia, Universidad Anáhuac Online, México, 2020; y también se puede consultar: "EXPERT SYSTEM

A partir de 1994, la tecnología de los SED se acopló a la recopilación de datos en línea y se abrió a una diversidad de modelos o metodologías que se están utilizando actualmente. S. Liao, hace una interesantísima clasificación de los SED que se utilizan actualmente en su texto: *Expert system methodologies and applications - a decade review from 1995 to 2004*, de la cual mostraremos solo los conceptos y aplicaciones principales:⁷⁷

- a) Sistemas basados en reglas (SED básico).- Las reglas son la base de las inferencias (IF-THEN). Las aplicaciones de los sistemas basados en reglas incluyen: análisis de transición estatal, tratamiento psiquiátrico, planificación de la producción, sistema de asesoramiento, enseñanza, planificación electrónica de energía, planificación de procesos automovilísticos, desarrollo de sistemas, verificación/validación de conocimientos, producción de alcohol, interpretación del histograma de ADN, mantenimiento de bases de conocimiento, evaluación de fraudes, adquisición de conocimiento, representación de conocimiento, diagnóstico de fallas del sistema de comunicación, diseño de procesamiento de materiales, utilización de recursos, nanotecnología bioquímica, diagnóstico de fallas bilísticas, planificación agrícola, apicultura, sistema de tutoría, geociencia y control de sensores.
- b) Sistemas basados en conocimiento. - Sus componentes son: una base de conocimiento, un motor de inferencia, una herramienta de Ingeniería de conocimiento y una interfaz de usuario específica. Incluye todas las

ENVIRONMENT/MVS" en el sitio: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?appid=skmwww&htmlfid=877%2FENUSZP86-3031&infotype=AN&subtype=CA&mhsrch=ibmsearch_a&mhq=EXPERT%20SYSTEM [20 de julio de 2022]

⁷⁷ Para conocer más detalles de este tema consúltese el interesantísimo estudio de S. Liao, (2005). *Expert system methodologies and applications - a decade review from 1995 to 2004*. *Expert Syst. Appl.*, 28, 93-103. <https://www.semanticscholar.org/paper/Expert-system-methodologies-and-applications-a-from-Liao/805530c2bf8ee064f49784679b9f8f4bdbc91620> [20 de julio de 2022] El autor hace una búsqueda de todos los sistemas expertos que existieron de 1995 a 2004 y los clasifica. Concluye que las metodologías de SED tienden a evolucionar hacia la orientación a la experiencia y que el desarrollo de aplicaciones de SED es un dominio orientado a problemas. Recomienda implementar las diferentes metodologías de las ciencias sociales, como la psicología, la ciencia cognitiva y el comportamiento humano. La clasificación que hace el autor es tomada para efectos de nuestras reflexiones. El texto marcado con negritas en la clasificación SED pretende destacar las áreas, actividades o aplicaciones utilizados en los servicios de salud.

aplicaciones de tecnología de la información organizacional que pueden resultar útiles para gestionar los activos de conocimiento de una organización, como los sistemas basados en reglas y de gestión de bases de datos. Las aplicaciones que utilizan los Sistemas basados en conocimiento son: tratamiento médico, planificación de finanzas personales, análisis de fallas de Ingeniería, gestión de residuos, gestión de la producción, Ingeniería térmica, apoyo a la toma de decisiones, gestión del conocimiento, representación del conocimiento, diseño de electrónica de potencia, evaluación de edificios enmarcados, análisis financiero, gestión de incidentes químicos, previsión climática, gestión agrícola, diseño de composición de acero, gestión estratégica, protección del medio ambiente, tratamiento de aguas residuales, aprendizaje, control de procesos químicos, planificación de terapia, planificación de localización de interrupciones, diseño de sistemas simultáneos, validación de casos, diseño de chips, planificación agrícola, protección transmisión de energía y diseño urbano.

- c) Redes neuronales artificiales⁷⁸. - Se utiliza para implementar simulaciones de software para procesos masivamente paralelos que involucran elementos de procesamiento interconectados en la arquitectura de red. La neurona artificial recibe entradas que son análogas a los impulsos electroquímicos que las dendritas de las neuronas biológicas reciben de otras neuronas. La salida de la neurona artificial corresponde a las señales enviadas desde una neurona biológica sobre su axón. Estas señales artificiales se pueden cambiar de manera similar a los cambios físicos que ocurren en las sinapsis neuronales. Las aplicaciones que utilizan Redes Neuronales son: diagnóstico de fallas, flujo de potencia óptimo, toma de decisiones, sistema de procesamiento de alarmas, mecanismos de inferencia, sistema de diagnóstico, aprendizaje automático, previsión de carga de potencia, diseño de instalaciones, control de procesos, aprendizaje de conocimientos, diseño de procesos de minería de oro, sistemas robóticos, configuración de

⁷⁸ Estas redes neuronales son el antecedente y base de las modernas “Redes Neuronales Convolucionales Profundas” que se utilizan en *Deep Learning*.

parámetros, tratamiento de residuos, control de procesos de mitigación, diagnóstico de señal acústica, destilación de aceite crudo y aplicación biomédica.

- d) Sistemas basados en lógica difusa. - El método de lógica difusa se ocupa de la “incertidumbre”, utilizando la teoría matemática de conjuntos difusos, simula el proceso de razonamiento humano normal al permitir que el ordenador se comporte con menos precisión y lógica que los ordenadores convencionales. Este enfoque se utiliza porque la toma de decisiones no siempre es una cuestión de blanco y negro, verdadero o falso; a menudo hay áreas grises y utilizamos el concepto “puede ser”. En consecuencia, los procesos creativos de toma de decisiones se pueden caracterizar como desestructurados, aleatorios, contenciosos y vagos. Las aplicaciones que utilizan Lógica difusa son: previsión de carga de energía, programación en línea, diagnóstico de fallas de proceso, planificación ecológica, control de sistemas, razonamiento incierto, integración de conocimiento, diagnóstico de fallas, clasificación del sistema de energía, detección de fallas, tratamiento de aguas residuales, previsión de suministro de agua, clasificación de radiografías, procesamiento analítico en línea, selección de hoteles, integración de herramientas de secadora, análisis de frecuencia de inundación agrupada, sistema de consulta médica, indexación de rendimiento, seguridad informática, reconocimiento de gestos.
- e) Sistemas basados en metodología orientada a objetos. - La metodología orientada a objetos combina en un objeto los datos junto con los procedimientos específicos que operan sobre estos datos, donde el objeto combina datos y código de programa. En lugar de pasar datos a los procedimientos, los programas envían un mensaje para que un objeto realice un procedimiento que ya está incrustado en él. Luego, el mismo mensaje puede enviarse a muchos objetos diferentes, pero cada uno implementará ese mensaje de manera diferente. Los datos de un objeto se encapsulan a partir de otras partes del sistema, por lo que cada objeto es un bloque de construcción de software independiente que se puede utilizar en muchos

sistemas diferentes sin cambiar el código del programa. Algunas aplicaciones implementadas por la metodología orientada a objetos incluyen las siguientes: diagnóstico de una industria, aprendizaje del conocimiento, red de información de fabricación, mantenimiento del sistema de energía, Ingeniería del conocimiento, programación sintáctica y representación del conocimiento.

- f) Sistemas de razonamiento basados en casos. - La idea básica está en adaptar soluciones que se utilizaron para resolver problemas anteriores y utilizarlos para resolver nuevos problemas. Las descripciones de la experiencia pasada de especialistas humanos, representados como casos, se almacenan en una base de datos para su posterior recuperación cuando el usuario encuentra un nuevo caso con parámetros similares. El sistema busca casos almacenados con características problemáticas similares a la nueva encomienda y encuentra el ajuste más cercano aplicando las soluciones del caso que le precede al nuevo caso. Las soluciones correctas se etiquetan en el nuevo caso y ambas se almacenan junto con los demás casos de la base de conocimiento. Las soluciones infructuosas también se añaden a la base de casos junto con explicaciones sobre por qué las soluciones no funcionaron. Algunas de las aplicaciones implementadas incluyen: diseño de procesos de fabricación, gestión del conocimiento, capacitación en restauración del sistema de energía, inspección ultrasónica, planificación médica, aplicación médica, diagnóstico de fallas, e-learning y modelado de conocimiento.
- g) Arquitectura de sistemas. - La arquitectura de sistemas parte de los bocetos que se utilizan en la arquitectura tradicional. Da a los usuarios una idea general de cómo se verá el sistema y cómo va a implementar los sistemas. La arquitectura muestra las capacidades generales del sistema, las interfaces de los usuarios, las funciones del sistema, el flujo del sistema (datos), la gestión del sistema, el protocolo necesario y el lenguaje de

programación específico, CommonKADS⁷⁹, etc. Los usuarios pueden manipular y controlar las funciones del sistema en la arquitectura del sistema. Algunas de las aplicaciones implementadas son: evaluación y selección de materiales, diseño asistido por computadora, diseño ergonómico, implementación del sistema ISO, soporte de decisiones de recuperación corporativa, Ingeniería concurrente, aplicación militar, simulador de entrenamiento, entre otros.

- h) Agentes inteligentes. - Un agente inteligente es un programa de computadora que ayuda a un usuario con las tareas rutinarias de la computadora. Varios de los nombres utilizados para describir los agentes inteligentes incluyen agentes de software, asistentes y agentes múltiples. Algunas de las aplicaciones implementadas son: sistemas de tutoría, análisis y diseño de sistemas, mantenimiento de servicios electrónicos, representación de conocimientos, sistemas adaptativos, control de contaminación del aire, diseño de arquitectura de edificios, apoyo a decisiones agrícolas, simulación de la industria, e Ingeniería del conocimiento en la plataforma de internet.
- i) Modelado. - La metodología de modelado se convierte en una metodología interdisciplinaria de SED para construir relaciones formales con el diseño de modelos lógicos en diferentes dominios de conocimiento/problemas. Además, la tecnología de modelado puede proporcionar métodos cuantitativos para analizar datos para representar o adquirir conocimiento experto con programación lógica inductiva o algoritmos para que la IA, la ciencia cognitiva y otros campos de investigación puedan tener plataformas más amplias para implementar tecnologías para el desarrollo de SED. Las aplicaciones implementadas son: control de procesos, análisis médico, toma de decisiones de gestión, evaluación de software, validación de sistemas médicos, planificación y simulación de tareas de montaje, diseño de

⁷⁹ CommonKADS es una metodología diseñada por investigadores de la Unión Europea para el análisis y la construcción de Sistemas Basados en Conocimiento, se pretendía crear un estándar para la Ingeniería del conocimiento y sistemas de conocimiento.

terminales de transporte, asignación de proyectos y clasificación de hiperplasia endometrial.

- j) Ontología⁸⁰. - Es una representación formal de un determinado dominio o área del conocimiento. Se utiliza como base de la comunicación entre los expertos en el dominio y los ingenieros del conocimiento. En consecuencia, se puede representar un modelo de tarea / dominio reutilizable y se genera un código de programa informático en esa ontología para la adquisición de conocimientos, la reutilización y el aprendizaje heurístico. Algunas aplicaciones implementadas por la ontología incluyen las siguientes: soporte de decisiones médicas, reutilización del conocimiento, control preventivo, evaluación del paisaje, adquisición de conocimiento y poda heurística del ajedrez.
- k) Metodología de bases de datos (DBMS). - Una base de datos es una colección de datos organizados para servir eficientemente a muchas aplicaciones centralizando los datos y minimizando los datos redundantes. Un DBMS es el software que permite a una organización centralizar los datos, administrarlos de manera eficiente y proporcionar acceso a los datos almacenados por programas de aplicación. Sin embargo, algunas bases de datos grandes hacen que la detección de conocimiento sea computacionalmente costosa porque algunos dominios o conocimientos en segundo plano, ocultos en la base de datos, pueden guiar y restringir la búsqueda de conocimientos importantes. Las metodologías modernas de la base de datos necesitan procesar grandes volúmenes de datos, múltiples jerarquías y diferentes formatos de datos. Las aplicaciones que la utilizan son: planificación de sistemas de energía, planificación geográfica, sistema de información geográfica, medicina tradicional china y sistemas expertos en salud.

⁸⁰ La ontología busca quitar las ambigüedades de los conceptos, a través de un sistema de vocabulario, que se utiliza para la categorización de conceptos, traducción automática, ortografía, filtrado de textos, segmentación de textos, reconocimiento de voz, gramática. Una ontología estaría formada por clases (conceptos del dominio), propiedades (relaciones-atributos), individuos y axiomas.

Las siguientes tablas muestran las ventajas y desventajas que existen por el uso de los Sistemas Expertos Decisionales y la ubicación de los Sistemas Expertos Decisionales dentro del campo de la Inteligencia Artificial.

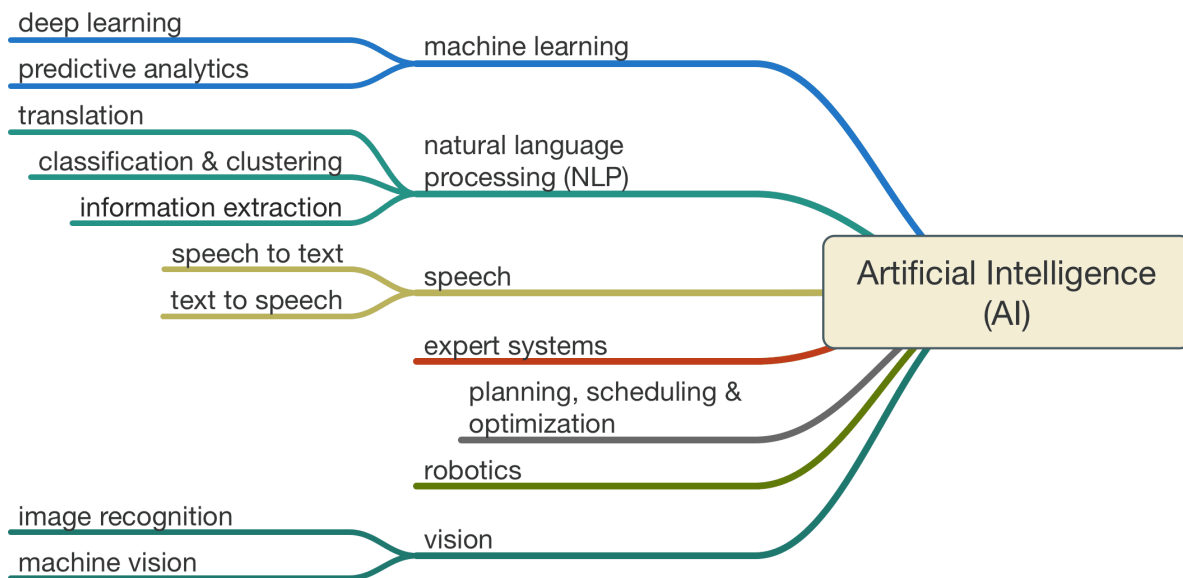
Ventajas y desventajas de los Sistemas Expertos Decisionales

Ventajas	Desventajas
• Incrementar la productividad de una industria • Incrementa la calidad reduciendo márgenes de error • Permitir la transmisión de experiencia y conocimientos • Mantener la objetividad en la aplicación de las normas (no corrupción) • Tener un grado alto de confiabilidad (seguridad jurídica) • Dar una vida ilimitada al vincularse con otros sistemas expertos y otras aplicaciones • Capacidad para procesar altos volúmenes de información • Obliga a la capacitación del personal y a su profesionalización • Reduce tiempos • Mejora la toma de decisiones	• Los avances tecnológicos pueden dejarlos limitados a determinados temas, por ejemplo, no tienen la capacidad de investigar algo nuevo. • Como otros experimentos de IA no tienen consciencia, no tienen sentido común. • Son solo un elemento de los muchos que debe tener una empresa. • No aprenden del experto, requieren de él para su actualización mediante el programador.

Fuente: El autor.

Para concluir esta sección, resta solamente reconocer que la mayor carga de producción de IA está en los SED. Se puede decir que son la columna vertebral de muchos sistemas y máquinas que se encuentran funcionando actualmente. La siguiente tabla ubica a los SED en la IA:

Ubicación de los Sistemas Expertos Decisionales dentro del campo de la Inteligencia Artificial



Fuente: Hill, Caroline, “Artificial Intelligence in Law – The State of Play in 2015?”, *Legal Technology*, 3 de noviembre de 2015. Consúltase en: <https://legaltechnology.com/artificial-intelligence-in-law-the-state-of-play-in-2015/> [15 julio de 2022]

En la siguiente sección analizaremos la IA basados en *Machine Learning – Deep Learning*, a través de la revisión de conceptos, evolución histórica, usos, tipos y características que los ubican como parte de la IA problemática no controlada por el ser humano que sí puede llegar a afectar Derechos Humanos.

Los sistemas expertos han tenido gran éxito para los servicios de salud y el trabajo del médico. Lo que se busca, como en cualquier actividad humana, es el perfeccionamiento y la profesionalización del sector de la salud en beneficio de las personas. Ya William Osler (1849-1919)⁸¹ decía que “La práctica de la medicina es un arte basado en la ciencia”⁸².

⁸¹ Osler vivió un momento histórico en el que la medicina evolucionó de una práctica basada en la superstición y la tradición a una ciencia biológica racional. Es considerado el padre de la medicina moderna y su libro formó a miles de generaciones de médicos. Cfr. William, Osler, *Principles and Practice of Medicine*, Appleton, New York, 1892.

⁸² Silverman M, Murray y TJ, Bryan C. *The Quotable Osler*, American College of Physicians, Filadelfia, 2003.

A partir de 1965, la mayoría de los programas de procesamiento de conocimiento médico se fueron construyendo bajo el modelo de los SED. Es decir, un programa que, utilizando conocimientos de un humano “experto en el tema o problema que se quiere atender”, emite una solución o recomendación que previamente fue ingresada por el programador con un formato basado -por lo general- en reglas, conocimiento o en otras metodologías.⁸³

En el campo de los servicios de la salud es muy frecuente que los médicos se encuentren ante decisiones complejas y que estas decisiones vayan acompañadas de un grado alto de “incertidumbre”, sobre todo cuando la vida del paciente está en juego. Los SED están hechos para la toma de decisiones y la reducción de esa incertidumbre. Las motivaciones para el desarrollo de SED han sido muchas, pero en conjunto, buscan ayudar a los médicos en el diagnóstico y tratamiento de las enfermedades que presentan sus pacientes.

Se parte de la idea de que un médico puede tener el conocimiento de la mayoría de las enfermedades, pero no con tanta precisión como los SED. Los beneficios que ofrecen son:

- a) Mejora la precisión del diagnóstico clínico a través de enfoques sistemáticos, completos y capaces de integrar datos de diversas fuentes.
- b) Aumenta la confiabilidad de las decisiones clínicas al evitar influencias injustificadas de casos similares, pero no idénticos.
- c) Mejora la rentabilidad de las pruebas y terapias al equilibrar los gastos de tiempo, las molestias contra los beneficios y los riesgos de las acciones definitivas.
- d) Aumenta nuestra comprensión de la estructura del conocimiento médico, con el desarrollo asociado de técnicas para identificar inconsistencias e insuficiencias en ese conocimiento.

⁸³ Puede consultarse, adicionalmente, a Edward A. Feigenbaum y Philip Klahr, “Expert Systems”. *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., Londres, pp. 684–689, en: <https://dl.acm.org/doi/10.5555/1074100.1074389> [20 de julio de 2022]

- e) Mejora nuestra comprensión de la toma de decisiones clínicas, a fin de mejorar la enseñanza médica y hacer que el sistema sea más efectivo y fácil de entender.⁸⁴

Hacer que un SED haga una búsqueda lo más exhaustiva posible para atender una problemática o para hacer una recomendación en la que el humano quedó limitado por el tiempo o por el volumen de información, es algo que se puede entender como una posibilidad adicional para incrementar el nivel de certeza en la toma de decisiones del personal de la salud. Además, dado que la naturaleza de un problema en medicina a veces no es posible de comprenderlo suficientemente, se sugiere utilizar los SED para analizar grandes cantidades de conocimiento y tomar sus resultados como un complemento de las decisiones del ser humano.

Un sistema experto debe usarse en la práctica médica sólo si mejora la calidad de la atención del paciente a un costo aceptable en tiempo o dinero, o si mantiene el estándar de atención existente a un costo reducido en tiempo o dinero. Se entiende que se mejoró la calidad de atención mediante uno o más de los siguientes supuestos:

- a) precisión diagnóstica mejorada;
- b) mejores resultados terapéuticos;
- c) un sentido mejorado del bienestar de los pacientes;
- d) acceso más fácil y rápido a la información del paciente a través de mejores sistemas de mantenimiento de registros; y
- e) una mejor representación de los hechos en los registros médicos y una mejor documentación de los motivos de las acciones de los médicos.⁸⁵

El uso de SED, en la toma de decisiones médicas, comenzó con la implementación de programas que realizaban análisis estadísticos o diagnósticos de determinadas enfermedades. Aceptaban un conjunto de hallazgos y seleccionaban una enfermedad de un conjunto fijo de enfermedades, utilizando

⁸⁴ Soni, S. R., Khunteta y Gupta, M., op. cit., 703-706.

⁸⁵ *Idem*.

métodos como el reconocimiento de patrones a través de funciones de lógica difusa, teoría de decisión bayesiana o las técnicas de árbol de decisiones.⁸⁶

En la década de 1970, los SED estaban enfocados a realizar diagnósticos y hacer recomendaciones de terapias. Sin embargo, ninguno estaba en uso clínico de rutina debido a la limitación de las técnicas de representación del conocimiento y la resistencia de la mayoría de los médicos. Durante la década de 1970, la mayoría de los sistemas expertos se desarrollaron en estaciones de trabajo potentes, utilizando lenguajes como LISP, PROLOG y OPS. Esto puso el desafío de desarrollar sistemas en manos de unos pocos elegidos que podían permitirse las plataformas y tuvieron la paciencia para aprender las complejidades de los idiomas disponibles.

Durante la década de 1980, la proliferación de computadoras personales y la introducción de herramientas para la construcción de los SED (denominadas *expert system shells* o intérprete de órdenes o intérprete de comandos) hizo que su uso se multiplicara exponencialmente; y en la década de 1990, la complejidad y el volumen del conocimiento médico aumentaron, encontrándose aproximadamente un total de 233 SED entre 1992 y 1996.⁸⁷

En el proceso de desarrollo de los SED para medicina, los problemas metodológicos básicos iniciaron como la representación del conocimiento y los mecanismos de inferencia para posteriormente cambiar hacia interfaces de usuario adecuadas, una relación de costo-beneficio satisfactorio y la adaptación de las funciones del sistema a las solicitudes del entorno clínico (década de 1990).⁸⁸

La adquisición y la representación del conocimiento fueron consideradas durante mucho tiempo como la principal limitación en el desarrollo de sistemas expertos en el campo de la medicina. La mayoría de las incidencias de problemas de adquisición de conocimiento reportados, involucraron problemas con la “calidad del conocimiento”. Gran parte de esto se debió a que los médicos tendían a comunicar un conocimiento superficial en lugar de la estructura de conocimiento

⁸⁶ *Idem.*

⁸⁷ *Idem.*

⁸⁸ *Idem.*

profundo o les resultaba difícil describir los procedimientos y las rutinas para los algoritmos. Además, se tenía el obstáculo de la adquisición de conocimiento al momento de aplicar algún examen médico de rutina ya que los pacientes dan respuestas muy subjetivas y pueden contradecirse, a veces repetidamente, al describir los síntomas.⁸⁹

Otro de los obstáculos que tuvieron fue que el diagnóstico médico se basa en una explicación causal de lo que le está sucediendo al paciente, y la terapia se basa en predicciones sobre cómo se puede modificar el proceso de la enfermedad. Los modelos de conocimiento en medicina son incompletos por estas dos razones, o, mejor dicho, son aproximados y omiten niveles de detalle. Esto mejorará con el uso del ML y de las redes neuronales profundas, pero no desaparece, razón por la cual se sigue cuestionando el uso de la tecnología para la toma de las decisiones médicas.⁹⁰

Las metodologías o modelos que utilizaron los SED para su funcionamiento y perfeccionamiento en el campo de la medicina son:

1. Razonamiento basado en reglas: El razonamiento basado en reglas es la estructura más tradicional y el modelo básico de los SED. Utiliza el conocimiento codificado en la regla de generación (si X... entonces Y). Las reglas generalmente tienen una parte de condición (antecedente) y una parte de acción (conclusión). Cada regla representa una de las unidades de conocimiento relacionadas con un campo experto. Muchas reglas relacionadas pueden corresponder a una cadena de inferencia, que deduce conclusiones útiles de varios hechos conocidos. Gracias a la experiencia previa en el desarrollo de sistemas basados en reglas, incluido MYCIN, se logró una mejor comprensión de su desarrollo, lo que provocó una explosión de aplicaciones. De los 260 sistemas de expertos médicos desarrollados desde 1992, 110 sistemas (42,3%) fueron desarrollados por el razonamiento basado en reglas. El razonamiento basado en reglas está controlado y tiene

⁸⁹ *Idem.*

⁹⁰ *Idem.*

beneficios naturales: Dado que las reglas son conocimientos independientes, pueden revisarse y verificarse fácilmente. Las reglas son transparentes, y ciertamente son mucho más transparentes que los modos de representación del conocimiento de tipo neuronal. Desventajas del enfoque de razonamiento basado en reglas: Requiere coincidencia exacta. Si no tenemos una coincidencia exacta, la regla no funcionaría. Las reglas no capturan de manera eficiente o natural la representación del conocimiento complejo. Tercero, los sistemas con un gran conjunto de reglas pueden ser lentos.

2. Razonamiento basado en casos. - Este enfoque resuelve problemas razonando desde principios. Es decir, explica el razonamiento al informar la serie de deducciones que condujeron desde los datos de entrada a la conclusión. Los médicos, sin embargo, resuelven nuevos problemas por analogía con casos antiguos y explican las razones en términos de experiencia previa. Los sistemas informáticos que resuelven por analogía se denominan sistemas de razonamiento basados en casos. Estos sistemas resuelven problemas buscando en una colección de casos almacenados para encontrar y recuperar los casos que más se asemejan (coinciden) a un caso recién presentado utilizando la puntuación de coincidencia. Un sistema de este tipo saca su poder de una gran biblioteca de casos, en lugar de un conjunto de principios. Los SED basados en casos se han aplicado a los siguientes temas médicos: análisis de la estructura del cráneo, cirugía general, radioterapia, bioquímica, atención postoperatoria, hemodinámica, atención de enfermería, psicoterapia, enfermedad cardíaca, procedimiento de diagnóstico por imágenes, análisis de proteínas, trastorno alimentario, rinitis alérgica, ictericia, VIH, radiología, función renal, asma y leucemia.
3. Enfoque estadístico bayesiano. - El enfoque bayesiano es un mecanismo para calcular la probabilidad de una enfermedad, tomando en cuenta la evidencia especificada, a partir de la probabilidad a priori de la enfermedad y las probabilidades condicionales que relacionan las observaciones con las enfermedades en las que pueden ocurrir. Se debe hacer una variedad de suposiciones adicionales en este enfoque, tales como: 1) Se supone que las

enfermedades bajo consideración son mutuamente excluyentes y exhaustivas (es decir, se supone que el paciente tiene exactamente una n cantidad de enfermedades); 2) Se supone que las observaciones clínicas son condicionalmente independientes sobre una enfermedad determinada; y 3) se supone que la incidencia de los síntomas de una enfermedad es estacionaria (es decir, el modelo generalmente no permite cambios en los patrones de la enfermedad con el tiempo). Se ha demostrado que tiene algunas limitaciones: 1. La suposición de independencia condicional de los síntomas generalmente no se aplica y puede conducir a una sustancial equivocación en ciertas configuraciones. 2. La suposición de exclusividad mutua y exhaustividad de las categorías de enfermedades suele ser falsa. En la práctica real, las categorías de enfermedades concurrentes y superpuestas son comunes. 3. En muchos dominios puede ser inexacto suponer que las probabilidades condicionales relevantes son estables en el tiempo. Además, las categorías y definiciones de diagnóstico cambian constantemente, al igual que las técnicas de observación de los médicos, invalidando los datos previamente acumulados. Actualmente se utiliza poco este enfoque con SED.

4. Red neuronal. - Para resolver los problemas anteriores, se pensó en la red neuronal y en la posibilidad de adquirir una base de conocimiento directamente de la base de datos médica. Se componen de muchos elementos de procesamiento simples e interconectados que interactúan dinámicamente entre sí para aprender o responder a la información en lugar de simplemente llevar a cabo pasos algorítmicos o instrucciones programadas. La información se representa en una red neuronal en ese patrón de fuerzas de interconexión entre los elementos de procesamiento. La información se procesa mediante un patrón cambiante de actividad distribuida en muchas unidades. Se ha aplicado un mecanismo de autoaprendizaje de la red neuronal a los sistemas médicos expertos como un método automático de adquisición y representación de conocimiento. Las redes neuronales tienen varias ventajas y desventajas en comparación con

el sistema basado en reglas. Los sistemas basados en reglas tienden a ser específicos de un dominio y funcionan extremadamente bien cuando los problemas están bien definidos, pero las redes neuronales tienen una amplia capacidad de respuesta basada en su capacidad de proporcionar una clasificación general de un conjunto de entradas. La implementación del sistema basado en reglas puede ser un proceso largo dependiendo del tamaño del dominio y el rango de casos que deben realizarse. Las redes neuronales pueden analizar una gran cantidad de casos rápidamente para proporcionar respuestas precisas. Sin embargo, la validación del contenido de la red neuronal es relativamente más difícil que el razonamiento basado en reglas. El uso de redes neuronales en la toma de decisiones médicas ha sido evaluado por su confiabilidad de la salida de las redes. La salida de los sistemas de redes neuronales es difícil de interpretar para la mayoría de los profesionales clínicos.⁹¹

De 1994 a 2005, los SED comenzaron a utilizar el internet y también comenzaron a incorporar más metodologías o modelos para mejorar su funcionamiento dentro de los servicios de salud. A las cuatro metodologías anteriores, que se pueden considerar como clásicas, se le suman otras que se encuentran en la siguiente tabla y que conjunta o separadamente son utilizadas hasta nuestros días:

⁹¹ *Idem.*

Sistemas Expertos Decisionales: Principales metodologías y aplicaciones para la salud⁹²

Metodología	Aplicaciones en medicina
Sistemas basados en reglas (SED básico)	• sistema de asesoramiento, enseñanza • desarrollo de sistemas • verificación/validación de conocimientos • interpretación del histograma de ADN • adquisición de conocimiento representación de conocimiento • nanotecnología bioquímica • diagnóstico de fallas bilísticas • sistema de tutoría • control de sensores
Sistemas basados en conocimiento	• tratamiento médico, • apoyo a la toma de decisiones • gestión del conocimiento • representación del conocimiento • aprendizaje • control de procesos químicos • planificación de terapia • validación de casos • diseño de chips
Redes neuronales artificiales⁹³	• toma de decisiones • mecanismos de inferencia • sistema de diagnóstico • aprendizaje automático • aprendizaje de conocimientos • configuración de parámetros • control de procesos de mitigación • aplicación biomédica
Sistemas basados en lógica difusa	• programación en línea • control de sistemas • razonamiento incierto • integración de conocimiento • diagnóstico de fallas • detección de fallas • clasificación de radiografías • procesamiento analítico en línea • sistema de consulta médica • seguridad informática • reconocimiento de gesto
Sistemas basados en metodología orientada a objetos	• aprendizaje del conocimiento • representación del conocimiento
Razonamiento basado en casos	• gestión del conocimiento

⁹² Adicional a este estudio, encontramos otro que puede servir de complemento, consúltense con la siguiente referencia: Haider Khalaf Jabbar and Rafiqul Zaman Khan. 2016. *Survey on Development of Expert System from 2010 to 2015. In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (ICTCS '16)*. Association for Computing Machinery, New York, NY, USA, Article 130, 1–7. DOI: <https://doi.org/10.1145/2905055.2905190> [20 de julio de 2022]

⁹³ Estas redes neuronales son el antecedente y base de las modernas “Redes Neuronales Convolucionales Profundas” que se utilizan en *Deep Learning*.

	• planificación médica • aplicación médica • diagnóstico de fallas • e-learning y modelado de conocimiento.
Arquitectura de sistemas	• implementación del sistema ISO • simulador de entrenamiento
Agentes inteligentes:	• sistemas de tutoría • análisis y diseño de sistemas • representación de conocimientos • sistemas adaptativos
Modelado	• control de procesos • análisis médico • toma de decisiones de gestión • evaluación de software • validación de sistemas médicos • clasificación de hiperplasia endometrial
Ontología⁹⁴:	• soporte de decisiones médicas • reutilización del conocimiento • control preventivo • adquisición de conocimiento
Metodología de bases de datos	• medicina tradicional china • sistemas expertos en salud

Fuente: El autor con base en: Liao, S., "Expert system methodologies and applications - a decade review from 1995 to 2004." en *Expert Syst. Appl.* No. 28, 2005, pp. 93-103. <https://www.semanticscholar.org/paper/Expert-system-methodologies-and-applications-a-from-Liao/805530c2bf8ee064f49784679b9f8f4bdbc91620> [15 julio de 2022]

Algunos de los SED más comentados en el mundo de la medicina son los siguientes:

1. DENDRAL, el primer sistema experto enfocado a la medicina, desarrollado por Edward Feigenbaum en la Universidad de Stanford. Facilitaba la inferencia en estructuras moleculares. Fue programado utilizando el lenguaje LISP.
2. MYCIN, es creado por Edward Shortliffe, en la Universidad de Stanford, se creó para diagnosticar enfermedades infecciosas de la sangre; este modelo ha servido para influir sustancialmente en gran parte del trabajo posterior en la construcción e implementación de sistemas expertos. MYCIN fue diseñado

⁹⁴ La ontología busca quitar las ambigüedades de los conceptos, a través de un sistema de vocabulario, que se utiliza para la categorización de conceptos, traducción automática, ortografía, filtrado de textos, segmentación de textos, reconocimiento de voz, gramática. Una ontología estaría formada por clases (conceptos del dominio), propiedades (relaciones-atributos), individuos y axiomas.

únicamente como un esfuerzo de investigación para demostrar cómo se pueden construir sistemas expertos para problemas del mundo real razonablemente grandes y complejos. La base de conocimientos de MYCIN contiene las reglas heurísticas utilizadas por los médicos en la identificación de ciertas infecciones.

3. PUFF se desarrolló en 1979, utilizando el shell EMYCIN. El propósito de PUFF es interpretar mediciones relacionadas con pruebas respiratorias e identificar trastornos pulmonares como por ejemplo el soplo. Al finalizar la prueba, PUFF presenta al médico su interpretación de las mediciones, un diagnóstico de la enfermedad y un esquema de tratamiento propuesto. La primera versión de PUFF tenía 64 reglas de producción. Una versión más reciente (codificada en BASIC) tenía alrededor de 400 reglas. El procedimiento de validación para PUFF se realizó comparando su diagnóstico con el de dos fisiólogos pulmonares expertos. La conclusión de PUFF y la de los fisiólogos fueron consistentes en más del 90 por ciento de los casos de prueba. Como resultado, PUFF ahora se usa de forma rutinaria. Una característica interesante particular de PUFF es que la apariencia física es indistinguible de las herramientas de laboratorio convencionales. La aceptación de PUFF, por la profesión médica, puede basarse en gran parte debido a esta percepción.
4. INTERNIST, se inició a principios de la década de 1970 y continuó posteriormente con el nombre de CADUCEUS. Tenía como objeto realizar un diagnóstico de las enfermedades asociadas con el campo de la medicina interna; tenía una base de conocimiento masiva.
5. QMR, utilizando la base de conocimiento masiva desarrollada por primera vez para INTERNIST, QMR (referencia médica rápida) ayuda a los médicos en el diagnóstico de una enfermedad basada en los síntomas del paciente, los resultados del examen y las pruebas de laboratorio. QMR, utilizado en la Universidad de Pittsburgh, incorpora más de 4.000 posibles manifestaciones de enfermedades y se dice que funciona a un nivel comparable al de los médicos en ejercicio.

6. Sistema de análisis de EEG, fue desarrollado para la detección automática de picos y ondas agudas para la detección de actividad epileptiforme. El sistema consta de dos etapas distintas. El primero es un extractor de características, escrito en FORTRAN, que utiliza algoritmos de detección de espigas para producir una lista de todas las ocurrencias similares a espigas en el EEG. La segunda etapa, escrita en OPS5 “shell”, lee la lista y usa reglas que incorporan el conocimiento obtenido para confirmar o excluir cada uno de los posibles picos.

Principales temas de los Sistemas Expertos Decisionales en los servicios de salud

<u>Beneficios</u>	<u>Metodologías utilizadas</u>	<u>Ejemplos de S.E.</u>
• Mejora la precisión del diagnóstico • Aumenta la confiabilidad de las decisiones clínicas • Mejora la rentabilidad de las pruebas y terapias al equilibrar los gastos de tiempo • Aumenta la comprensión de la estructura del conocimiento médico • Mejora la comprensión de la toma de decisiones clínicas	• Enfoque estadístico bayesiano • Razonamiento basado en reglas • Red neuronal • Razonamiento basado en casos	INTERNIST MYCIN PUFF QMR EGG
Discusiones respecto su uso • Problemas éticos y legales • Problemas de comportamiento • Errores en la implementación de metodología		
1960 Programas que realizaban análisis estadísticos; centrados en la parte de diagnóstico /reconocimiento de patrones (teoría bayesiana, árbol de decisiones, funciones discriminatorias)	1970 Diagnósticos y recomendaciones de terapias. No estaba en usos clínicos de rutina	1980-1990 Proliferación de S.E. por las PC y de herramientas de software “shells”

Fuente: Elaboración propia.

Para finalizar esta sección, mostraremos algunos ejemplos recientes de investigaciones sobre SED que se han hecho de 2016 a 2020 con el fin de fomentar su uso:

- eRheumatologist: sistema experto basado en dispositivos móviles para reumatología que utiliza un algoritmo de lógica difusa

Los autores del estudio⁹⁵ comentan que las afecciones reumáticas o musculoesqueléticas comprenden más de 150 enfermedades y síndromes, que suelen ser progresivos y asociados con el dolor. Existen enfermedades articulares, discapacidad física, trastornos espinales y afecciones resultantes de traumatismos. Las condiciones musculoesqueléticas son las principales causas de morbilidad y discapacidad, lo que da lugar a enormes gastos sanitarios y pérdida de trabajo. Se desarrolló un SED basado en telefonía móvil enfocado a la enfermedad reumática que permite a los usuarios diagnosticar su condición actual. Permite al usuario identificar la posible causa de sus articulaciones y dolor muscular. El sistema aplica el concepto de algoritmo lógico difuso para poder obtener el posible resultado de las pruebas diagnósticas. Este algoritmo permitirá al sistema determinar y categorizar posibles síntomas y evaluado por las diferentes reglas como parte del motor de inferencia y validado por la base de datos basada en el conocimiento. El sistema fue evaluado por diferentes usuarios como un reumatólogo, un desarrollador móvil y usuarios de diferentes edades para garantizar que el sistema alcance sus objetivos. Como resultado de la evaluación, los usuarios estuvieron de acuerdo en que la aplicación realizaba sus funciones como una herramienta de diagnóstico. El objetivo es implementar un algoritmo de lógica difusa para un grupo con perfil de enfermedades reumáticas, desarrollar la aplicación que ayude al monitoreo de usuarios y evaluar la aplicación sobre su funcionalidad y procesamiento de la información (en sistemas Android e iOS). El algoritmo de lógica difusa permite al SED determinar y categorizar los posibles síntomas y evaluar las diferentes reglas como parte del motor de inferencia y es validado por la base de datos en el conocimiento del experto médico.

⁹⁵ Mark Jade S. Corpuz, Junnel Avestro, Fredilyn B. Calanda, Roselia C. Morco, Jean M. Ángeles, and Carl Matthew S. Alcala, "Rheumatologist: Mobile-based Expert System for Rheumatology Utilizing Fuzzy Logic Algorithm", en *Proceedings of the 2017 International Conference on Computer Science and Artificial Intelligence (CSAI 2017)*, Association for Computing Machinery, New York, 2017, pp. 153–157. DOI: <https://doi.org/10.1145/3168390.3168428> [20 de julio de 2022]

- Un marco para diseñar sistemas exitosos de apoyo a la toma de decisiones clínicas

Los autores del estudio⁹⁶ desarrollan siete principios para el modelado exitoso del proceso clínico, formando un marco para el diseño de sistemas de apoyo a la toma de decisiones clínicas. Los métodos de modelado -comentan- deben incorporar interacciones de datos durante las decisiones clínicas y deben imitar las habilidades cognitivas de los médicos. Los modelos predictivos deben ser interactivos, regenerando predicciones en respuesta a nueva información clínica o comentarios de los médicos. Cualquier método de apoyo a la decisión debe considerar las tendencias de las mediciones fisiológicas. Las tendencias temporales pueden ser predictores más fuertes de los resultados de salud que los valores transversales. Por último, los métodos de apoyo a las decisiones clínicas deben basarse en los resultados, en un esfuerzo por evitar un sesgo de decisión histórica. Estos principios pueden contribuir a optimizar las metodologías de modelado en entornos sanitarios, mejorando la respuesta de los sistemas de salud a los desafíos de la toma de decisiones. Los principios se pueden encontrar en la tabla siguiente:

⁹⁶ Dimitrios Zikos. "A Framework to Design Successful Clinical Decision Support Systems", en *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '17)*, Association for Computing Machinery, New York, 2017, pp. 185–188. DOI:<https://doi.org/10.1145/3056540.3064960> [20 de julio de 2022]

Table 6: Significance of each principle and CDSS design considerations

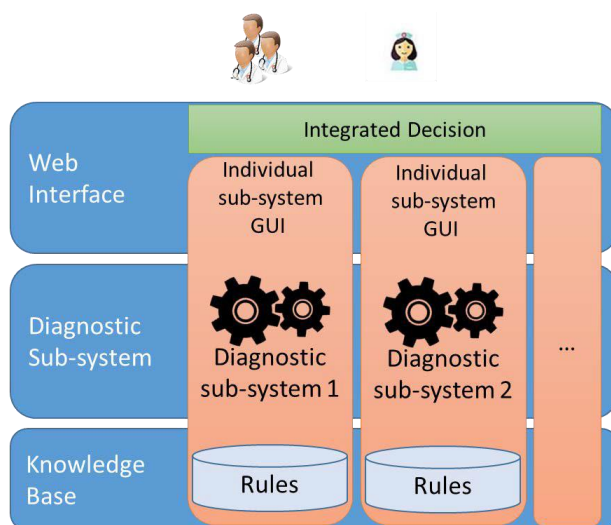
Principle	Significance	CDS system considerations
Principle 1: CDSS should mimic the cognitive process of clinical decision makers	Health data become useful when combined with human knowledge & experience	(1) Expert systems can be harmonically combined with machine learning (2) Interactive models considering new info & feedback from clinicians
Principle 2: CDSS should provide recommendations with longitudinal insight	Clinicians look for changes over time rather than raw measurement values	Models need to include, as predictors, trends of repeated measurements
Principle 3: CDS systems should 'know' the time decisions will be made	Data availability varies for different decision points	CDSS explicitly informing the user about the available data required for predictions
Principle 4: CDSS should provide dynamic predictions	During the hospital stay, for a new prediction, more information is available	Newly trained and tested models, on the fly
Principle 5: The 'Historical Decision' bias should be avoided	Mimicking wrong decisions of past datasets is not a good practice	Outcomes based designs approaches
Principle 6: CDSS should model known interactions between clinical variables	When combined, health data become useful for decisions	Models annotate a-priori known sets of variables that critical, altogether, as predictors
Principle 7: Dimensionality reduction allows on-the-fly training but should be used with caution	Increase model efficiency for dynamic predictions	(1) Principal Component Analysis methods (2) Regrouping ICD-10-CM diagnoses

El objetivo de la investigación está en ubicar algunas consideraciones importantes para el diseño de modelos y métodos predictivos para la toma de decisiones clínicas. En primer lugar, cualquier método debe tener en cuenta la naturaleza única de los datos en el sector salud. El sistema que soporte las decisiones médicas debe proporcionar las recomendaciones basadas en la evidencia a los médicos para mejorar su capacidad y sus conocimientos, para lograr un servicio seguro y de alta calidad. Destaca de nueva cuenta la mezcla de SED tradicionales con modelos predictivos de *Machine Learning*.

- INTEGRA: un sistema de diagnóstico diferencial basado en la web que combina múltiples bases de conocimientos

Los autores del estudio⁹⁷ comentan que INTEGRA permite hacer diagnósticos respiratorios para apoyo a la decisión médica y obtener una explicación estos resultados con base en el diagnóstico combinado de metodologías de SED. INTEGRA emplea una arquitectura de tres capas, consiste en la interfaz web, el subsistema de diagnóstico y la capa que contiene la base de conocimiento para el razonamiento. Puede trabajar con una base del conocimiento ya disponible por el usuario, con varias bases o con su propia base de conocimiento.

Arquitectura de INTEGRA



Fuente: INTEGRA: a web-based differential diagnosis system combining multiple knowledge bases

La capa de interfaz del usuario utiliza Javascript y HTML. El experto puede introducir las quejas del paciente y seleccionar los síntomas de una lista generada. Tan pronto como el experto hace las selecciones apropiadas, los resultados del diagnóstico se presentan junto con una probabilidad de confianza. En este estudio los autores consideran que el uso de las ontologías sirve como una metodología correcta para su utilización en los sistemas

⁹⁷ Cfr. Aris Papakonstantinou, Haridimos Kondylakis, y Emmanouil Marakakis, "INTEGRA: a web-based differential diagnosis system combining multiple knowledge bases", *In Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '20)*, Association for Computing Machinery, New York, 2020, article 45, pp. 1-6. DOI: <https://doi.org/10.1145/3389189.3397980> [20 de julio de 2022]

médicos y de salud de todo el mundo. Adicionalmente, comentan que existen otros SED para la toma de decisiones sobre procedimientos de diagnóstico médicos, como, por ejemplo: APACHE III, un sistema que intenta predecir el riesgo de que la persona muera en un hospital. Esta predicción se basa en una comparación de la historia clínica de 18.000 casos, almacenados en la base de datos del sistema. Tiene un promedio de 95% de precisión predictiva, y utiliza un mecanismo basado en partituras. También existe LISA. Un Sistema de Información Clínica y Apoyo a la Toma de Decisiones para la atención cooperativa de leucemia linfoblástica aguda infantil que se ocupa de proporcionar apoyo durante el período de tratamiento del paciente, donde se deben tomar decisiones semanales sobre la dosificación de fármacos. Otro sistema es “Isabel” es un sistema de apoyo a la toma de decisiones clínicas basado en la web que proporciona apoyo en decisiones diagnósticas pediátricas. Isabel utiliza el software de procesamiento de lenguaje natural de *Autonomy* y consiste en una base de datos médica patentada con más de 11,000 diagnósticos y 4,000 medicamentos. *Therapy Edge HIV* es un sistema de apoyo clínico de iones basado en el web e introducido en 2005 que se ocupa del tratamiento del VIH, su razonamiento se basa en directrices temporales para evaluar el estado actual del paciente y crear opciones de tratamiento alternativas.⁹⁸

2) Inteligencia Artificial basada en Machine Learning - Deep Learning.

La Inteligencia Artificial basada en *Machine Learning - Deep Learning*, la podemos ubicar pieza clave para el inicio de la llamada Inteligencia Artificial “fuerte o generalizada”, la cual puede interactuar y operar una variedad de tareas independientes y no relacionadas. Puede aprender por sí misma nuevas tareas y resolver nuevos problemas con nuevas estrategias. La IA fuerte o generalizada es la combinación de modelos algorítmicos que aprenden de la experiencia y pueden resolver o tomar decisiones parecidas a la que tomaría un ser humano.⁹⁹ Algunos sistemas tienen la posibilidad de tomar decisiones sin requerir de un programador ni de experto que la guíe, genera su propio conocimiento de los datos o del ambiente.

⁹⁸ *Idem.*

⁹⁹ Cfr. Rusell y Norvig, *op. cit.*, p. 26.

El sueño de los padres de la Inteligencia Artificial (IA) era lograr que las máquinas fueran tan inteligentes como los seres humanos. Si bien es cierto que seguimos en la construcción de esto, también lo es que la inteligencia artificial ahora trabaja con varios ingredientes que los científicos del siglo XX no tenían y que son los que dan soporte a este tipo de tecnología. Por ejemplo, hoy estamos trabajando con máquinas con gran capacidad de memoria, con procesadores más veloces y un almacenamiento inmensamente superior. La posibilidad de que las computadoras tengan mayor capacidad para procesar estos datos ha sido la clave para lograr muchos avances. Mientras que los procesadores de la década de 1970 podían ejecutar “92.000 instrucciones por segundo, los procesadores de los teléfonos inteligentes de hoy pueden ejecutar miles de millones de instrucciones por segundo. Siguiendo lo que se ha llamado la Ley de Moore, la capacidad de procesamiento de las computadoras ha aumentado enormemente en las últimas décadas, aproximadamente el doble cada dos años.”¹⁰⁰

Esto llevará a la ciencia al desarrollo de los inicios de una IA generalizada para la toma de decisiones sin ayuda de los seres humanos. En principio, diremos que existen tres momentos paradigmáticos en los que se busca que las máquinas tomen decisiones tal como lo hacen los seres humanos y que son considerados ejemplos de la IA fuerte o generalizada. El primer intento lo encontramos en Deep Blue (1997), que se convirtió en el primer sistema informático de juego de ajedrez que venció a un campeón mundial (Garry Kasparov)¹⁰¹. En ese momento, no se pensaba tanto en buscar nuevas opciones de modelos algorítmicos, sino en incrementar la potencia informática disponible en la década de 1990 para realizar búsquedas a gran escala de movimientos potenciales podía procesar más de 200 millones de movimientos por segundo, y luego elegir el mejor. Si bien, todavía está presente la actividad del programador humano, representaba un gran avance.

¹⁰⁰ Cfr. *Más de 50 años de la Ley de Moore*, Intel, consúltese en: <https://www.intel.la/content/www/xl/es/silicon-innovations/moores-law-technology.html> [20 de julio de 2022]

¹⁰¹ Un par de documentales sobre esta hazaña son: “El hombre contra la máquina” en <https://www.youtube.com/watch?v=JodijRqCg6k> y “Garry Kasparov versus Deep Thought Documentary” en <https://www.youtube.com/watch?v=ke8pq-cpOGk> [15 de julio de 2022] Kasparov ya había jugado partidas con máquinas en 1989 (Deep Thought) y en 1996 (Deep Blue, en su primera versión).

El segundo intento está en Watson de IBM (2011), se enfrentó a jugadores humanos en el concurso estadounidense Jeopardy y venció a dos de los campeones del programa, Brad Rutter y Ken Jennings. Estos sistemas basaron su éxito en reglas, que codificaban el conocimiento humano sobre cómo jugar y qué movimientos usar en diferentes situaciones. Las computadoras podían utilizar métodos de árbol de decisiones o búsqueda avanzada para seleccionar una respuesta apropiada para una configuración particular de piezas y usaban mayor potencia informática para realizar búsquedas complejas con el fin de seleccionar su próximo movimiento.

Hasta aquí vale la pena hacer una aclaración. Tanto en el primer intento como en el segundo intento, existían limitantes en la escalabilidad y en transferibilidad de los sistemas. Un sistema que competía en el juego de ajedrez no podía jugar a las damas, y un sistema que se basaba en este tipo de reglas no se podía escalar a juegos más desafiantes o intuitivos. En cambio, en el tercer intento encontramos los “inicios” de la llamada Inteligencia Artificial fuerte o generalizada.

¿Cómo surge ese tercer intento? El tercer intento lo encontramos con *Google DeepMind* (2015-2020) y se fue construyendo en varios momentos. El primero se ubica con una serie de partidos del juego Go¹⁰², y varios campeones humanos que participan en competencia contra *AlphaGo*. Uno de ellos fue Lee Sedol, quien era reconocido como el mejor jugador de Go del mundo. *AlphaGo* jugó cinco partidos contra Lee Sedol, ganando cuatro de ellos. Estas victorias demostraron la capacidad del aprendizaje automático para abordar tareas complejas y, al hacerlo, producir soluciones que los humanos hacen con su intuición.

En un segundo momento, los investigadores de *DeepMind*, cambiaron el enfoque y utilizaron ahora búsquedas estocásticas y redes neuronales profundas, entrenando a *AlphaGo* en 30 millones de movimientos de juegos hechos por humanos. Luego utilizaron el aprendizaje por refuerzo para permitir que *AlphaGo* aprendiera de miles de juegos que jugó contra sí mismo.

¹⁰² El juego Go se originó en China hace más de 2,500 años, los jugadores colocan piedras en un tablero y luchan con el objetivo de lograr controlar un mayor número de espacios (territorio) capturando las piedras de su oponente.

Para 2017, *DeepMind* nuevamente da otro salto creando *AlphaGo Zero*, un sistema que se enseñó a sí mismo (desde cero) a dominar los juegos de ajedrez, *Shogi* y *Go*, superando a otros programas informáticos campeones del mundo en cada caso. AlphaZero reemplaza la heurística hecha a mano con una red neuronal profunda y algoritmos que no reciben nada más allá de las reglas básicas del juego. Al enseñarse a sí mismo, *AlphaZero* desarrolló su propio estilo de juego.¹⁰³

Para 2019, el algoritmo de *DeepMind*, conocido como *MuZero*¹⁰⁴, utiliza ahora el rendimiento de *AlphaZero* para *Go*, Ajedrez y *Shogi* y al mismo tiempo domina toda una variedad de juegos de Atari visualmente complejos, todo sin que se le digan las reglas de ningún juego. Para ello, aprende de su entorno y lo combina con el árbol de anticipación de *AlphaZero*. Esto le permite planificar estrategias ganadoras en dominios desconocidos, un avance significativo en las capacidades de los algoritmos de aprendizaje por refuerzo y un paso importante para construir sistemas de aprendizaje de propósito general. La idea de los creadores de *MuZero* es allanar el camino para abordar problemas de entornos desordenados o complejos como el mundo real en donde se desconocen las reglas del juego.

Para el uso del Aprendizaje Profundo Reforzado o Aprendizaje por Refuerzo Profundo (*Deep Reinforcement Learning*), en ajedrez, por ejemplo, los científicos se han basado en jugadores expertos de los que surgieron casi todas las mejoras del sistema. Matthew Lai explica la relevancia de este tipo de modelos algorítmicos concluyendo que el llamado algoritmo Giraffe fue de los primeros en ser capaz de aprender del auto-juego, y del resultado del entrenamiento en tan sólo 72 horas en una máquina:

En este proyecto, investigamos el uso del **Aprendizaje por Refuerzo Profundo** con extracción automática de características en el juego de ajedrez. Los resultados muestran que el sistema funciona al menos de manera comparable a los mejores expertos humanos de la actualidad, muchas de las jugadas se han perfeccionado a lo largo de décadas. La belleza de este enfoque está en su generalidad... es probable que este enfoque pueda trasladarse fácilmente a otros juegos de mesa de suma cero basados en turnos y lograr un rendimiento de vanguardia rápidamente... para crear un jugador fuerte de IA. Además de los

¹⁰³Una explicación general se puede consultar en: *A Simple Alpha (Go) Zero Tutorial*, 29 de diciembre de 2017, Stanford <https://web.stanford.edu/~surag/posts/alphazero.html> [20 de julio de 2022]

¹⁰⁴ Cfr. J., Schrittwieser, I., Antonoglou, T., Hubert, *et al.*, "Mastering Atari, Go, chess and shogi by planning with a learned model", *Nature*; número 588, 2020, pp. 604-609. <https://doi.org/10.1038/s41586-020-03051-4> [La traducción es propia] [20 de julio de 2022]

aspectos de aprendizaje automático del proyecto, presentamos y probamos una variante alternativa del **algoritmo minimax** de décadas de antigüedad, donde aplicamos límites de probabilidad en lugar de límites de profundidad para limitar el árbol de búsqueda. Demostramos que este enfoque es al menos comparable y posiblemente superior al enfoque que se ha utilizado durante el último medio siglo... **La eficiencia es siempre una consideración importante cuando se cambia de un sistema experto a un enfoque de aprendizaje automático**, ya que los sistemas expertos suelen ser más eficientes de evaluar que los modelos genéricos... Algunos intentos anteriores de aplicar redes neuronales al ajedrez se han visto frustrados por grandes penalizaciones de rendimiento. La implementación optimizada de la red neuronal de Giraffe, cuando se combina con el rendimiento aritmético vectorial lo convierte en un jugador bastante más competitivo en comparación con muchos motores de ajedrez. Con todas nuestras mejoras, Giraffe puede jugar al nivel de un maestro internacional de la FIDE en una PC convencional moderna. Uno de los objetivos originales del proyecto es crear un motor de ajedrez que dependa menos de la fuerza bruta que sus contemporáneos, y ese objetivo ciertamente se ha logrado. A diferencia de la mayoría de los motores de ajedrez que existen en la actualidad, Giraffe obtiene su fuerza de juego no de poder ver muy lejos, sino de poder evaluar posiciones difíciles con precisión y comprender conceptos posicionales complicados que son intuitivos para los humanos, pero que han sido difíciles de alcanzar... Esto es especialmente importante en las fases de apertura y final del juego, donde se juega excepcionalmente bien. ¹⁰⁵

Construir sistemas autónomos con capacidades de planificación ha sido durante mucho tiempo uno de los principales desafíos de la inteligencia artificial. Sin embargo, en los problemas del mundo real, las dinámicas que gobiernan el medio ambiente son típicamente complejas y desconocidas, y los enfoques de planificación hasta ahora no han logrado obtener las mismas ganancias en el desempeño.

A finales de 2020, nuevamente evoluciona el algoritmo *MuZero*. El algoritmo *MuZero*, ha igualado el rendimiento sobrehumano de los algoritmos de planificación de alto rendimiento en sus dominios preferidos (juegos de mesa lógicamente complejos como el ajedrez y el Go) y ha superado a los algoritmos RL sin modelos de última generación en sus dominios preferidos (juegos Atari visualmente complejos). El algoritmo Mu Zero no requiere ningún conocimiento de la dinámica del entorno, potenciando el camino hacia la aplicación de potentes métodos de aprendizaje y planificación a una serie de dominios del mundo real para los que no

¹⁰⁵ Matthew, Lai, "Giraffe: Using Deep Reinforcement Learning to Play Chess", ArXiv abs/1509.01549, 2015, s/p.

existe un simulador perfecto. La revista *Nature* ha publicado un interesante artículo sobre la evolución de este algoritmo:

“El algoritmo MuZero aprende un modelo iterable que produce predicciones relevantes para la planificación. Se probó el algoritmo en 57 juegos de Atari diferentes logrando un rendimiento de vanguardia. Cuando se evalúa en Go, en ajedrez y en shogi coincidieron, sin ningún conocimiento de la dinámica del juego, con el rendimiento sobrehumano del algoritmo AlphaZero que se suministraba con las reglas del juego. Los algoritmos de planificación han logrado notables éxitos en inteligencia artificial. Los campeones del mundo han sido derrotados en juegos clásicos como damas, ajedrez, Go y Poker y los algoritmos de planificación han tenido impacto en el mundo real en aplicaciones desde la logística hasta la síntesis química. Sin embargo, todos estos algoritmos de planificación se basan en el conocimiento de la dinámica del entorno, como las reglas del juego o un simulador preciso, impidiendo su aplicación directa a dominios del mundo real como la robótica, el control industrial o los asistentes inteligentes, donde la dinámica es normalmente desconocida. El aprendizaje por refuerzo (RL, *Reinforcement Learning*) tiene como objetivo abordar este problema aprendiendo primero un modelo de la dinámica del entorno y luego planificando con respecto al modelo aprendido. Típicamente, estos modelos se han centrado en la reconstrucción del verdadero estado ambiental o la secuencia de observaciones completas. Sin embargo, el trabajo anterior permanece lejos del estado de la técnica en dominios visualmente ricos. En su lugar, los métodos más exitosos se basan en RL sin modelos, es decir, estiman la política óptima y/o función de valor directamente a partir de interacciones con el entorno. Aquí presentamos MuZero, un nuevo enfoque de RL basado en modelos que logra tanto un rendimiento de vanguardia en los juegos Atari 2600, un conjunto visualmente complejo de dominios, como un rendimiento sobrehumano en tareas de planificación de precisión como ajedrez, Shogi y Go, sin conocimiento previo de la dinámica del juego. MuZero se basa en los potentes algoritmos de búsqueda e iteración de políticas de AlphaZero, pero incorpora un modelo aprendido en el procedimiento de entrenamiento. MuZero también extiende AlphaZero a un conjunto más amplio de entornos, incluidos dominios de agente único y recompensas distintas de cero en pasos de tiempo intermedios. La idea principal del algoritmo es predecir aquellos aspectos del futuro que son directamente relevantes para la planificación. El modelo recibe la observación (por ejemplo, una imagen de la placa Go o la pantalla Atari) como entrada y la transforma en un estado oculto. En cada uno de estos pasos, el modelo produce una política (predecir el movimiento a juego), la función de valor (predecir la recompensa acumulada, por ejemplo, el ganador final) y la predicción inmediata de recompensas (por ejemplo, los puntos anotados al jugar un movimiento). El modelo se entrena de extremo a extremo, con el único objetivo de estimar con precisión estas tres cantidades importantes, para que coincida con la mejora de la política y la función de valor generada por la búsqueda, así como la recompensa observada.”¹⁰⁶

El algoritmo entrenado superó el rendimiento de otros sistemas de IA y el rendimiento de expertos humanos. “El modelo es una red neuronal convolucional,

¹⁰⁶ Schrittwieser, J., Antonoglou, I., Hubert, T. et al., “Mastering Atari, Go, chess and shogi by planning with a learned model”, *Nature*, número 588, 2020, pp. 604-609. <https://doi.org/10.1038/s41586-020-03051-4> [20 de julio de 2022]

entrenada con una variante de *Q-learning*, cuya entrada son píxeles sin procesar y cuya salida es una función de valor que estima recompensas futuras.”¹⁰⁷ Las siguientes tablas sintetizan esta evolución de la IA:

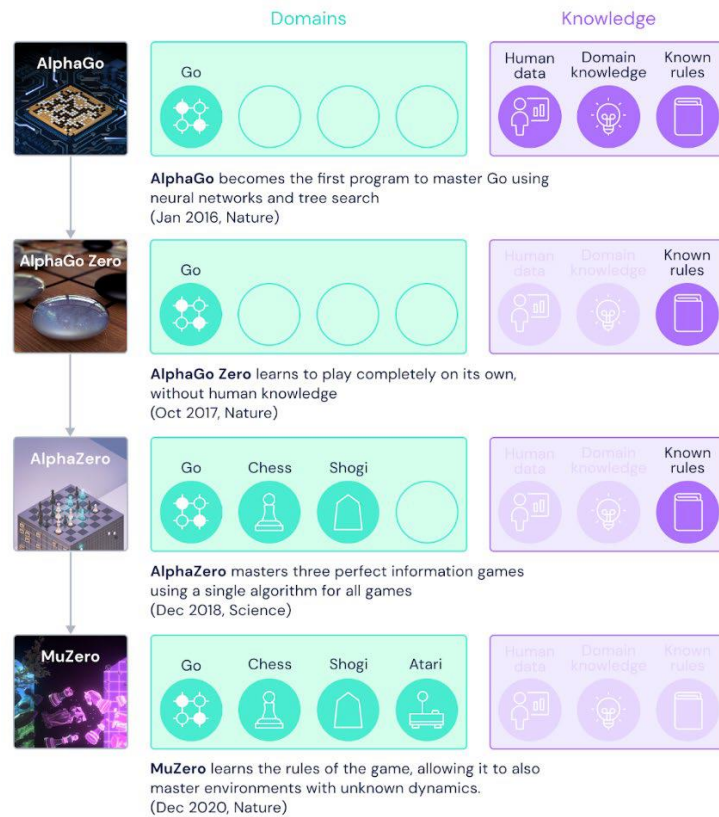
DeepMind (evolución de sus algoritmos)

Fecha	Azaña
Octubre de 2015	AlphaGo Fan derrotó a la campeona europea Fan Hui.
Enero de 2016	AlphaGo se convierte en el primer programa en dominar Go utilizando redes neuronales y búsqueda de árboles (Nature)
Marzo de 2016	AlphaGo Lee derrotó al jugador de World's Top Go, Lee Sedol, quien había ganado el campeonato mundial 18 veces.
Mayo de 2017	AlphaGo Master participó en la cumbre Future of Go. Ganó 60 juegos consecutivos en línea, eliminó a Ke Jie en un partido de tres juegos y venció a un equipo humano con cinco de los mejores profesionales de Go del mundo.
Octubre de 2017	AlphaGo Zero aprende a jugar completamente solo, sin conocimiento humano (Naturaleza)
Diciembre de 2017	Después de solo 40 días de auto-entrenamiento. AlphaGo Zero se hizo aún más fuerte, superando a AlphaGo Master, que ha derrotado a los mejores jugadores del mundo y al número uno del mundo, Ke Jie.
Diciembre de 2018	AlphaZero domina tres juegos de información perfecta usando un solo algoritmo para todos los juegos (Ciencia)
Diciembre de 2020	MuZero aprende las reglas del juego, lo que le permite también dominar entornos con dinámicas desconocidas. Octubre de 2017: después de solo tres días de aprendizaje reforzado (es decir, la computadora solo juega contra sí misma), AlphaGo Zero derrotó enfáticamente a AlphaGo Lee por 100 juegos a 0.

Fuente: El autor, con base en información de <https://deepmind.com/blog/article/muzero-mastering-go-chess-shogi-and-atari-without-rules> y <https://ndiastorage.blob.core.usgovcloudapi.net/ndia/2019/solic/Shaneman.pdf> [15 julio de 2022]

¹⁰⁷ Cfr. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M.A. “Playing Atari with Deep Reinforcement Learning”, *ArXiv, abs/1312.5602*, 2013.

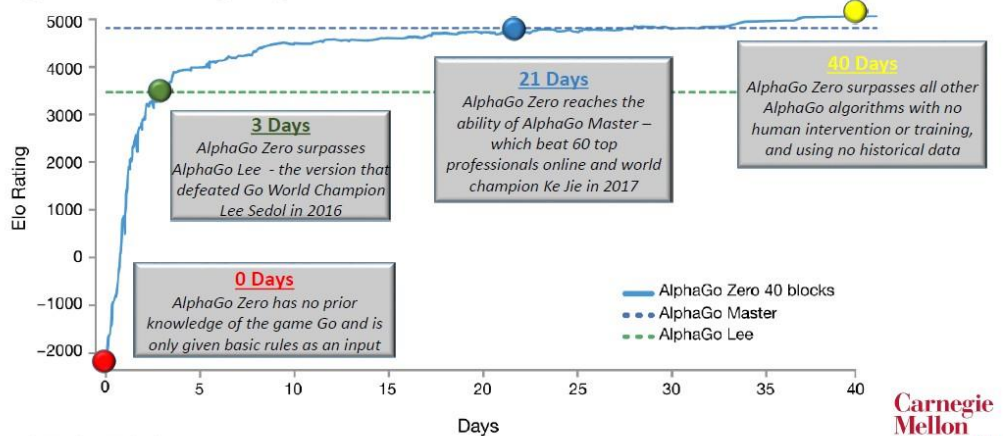
MuZero: Dominando Go, ajedrez, shogi y Atari sin reglas



Fuente: DeepMind, <https://deepmind.com/blog/article/muzero-mastering-go-chess-shogi-and-atari-without-rules> [15 julio de 2022]

Tiempos de aprendizaje de los algoritmos de DeepMind

Algorithmic Agility



Carnegie
Mellon
University

Para entender el funcionamiento de esta tecnología, se describirá en la siguiente sección la evolución histórica, tipos y principales usos de los algoritmos de “ML - DL” y la llegada de las Redes Neuronales Convolucionales Profundas.

A) Machine Learning

Entender la ruta evolutiva de este tipo de tecnología nos lleva a ubicar el surgimiento, histórico del “*Machine Learning*” y *Deep Learning*” (ML-DL). Ambos sumados a las nuevas tecnologías¹⁰⁸ están mejorando a gran velocidad y con mayor precisión diversas actividades de la empresa, el gobierno y la sociedad en conjunto. Pero a la vez, está generando afectaciones a Derechos Humanos.

Se le adjudica la invención del ML a Arthur Samuel. El salto a la fama de Samuel fue cuando en 1950 escribió un programa para jugar damas, y lo increíble fue que el mismo Arthur Samuel no era un buen jugador de damas y el programa logró resultados muy avanzados para esa época. Lo que hizo Samuel fue programar miles de partidas contra sí mismo, y ver qué posiciones en el tablero tendían a conducir a victorias y qué posiciones en el tablero tendían a conducir a derrotas.

El programa de Samuel utilizó un modelo de árbol binario de búsqueda para calcular los posibles movimientos y evaluar la posición del tablero. La máquina adquirió una comprensión de los movimientos buenos y malos a través de juegos repetidos y utilizó esto para jugar (aunque nunca logró un juego de nivel experto, se caracterizó por ser mejor que el promedio).

En ese momento se vio también, cómo una computadora podría desarrollar miles de partidas del juego de damas sin cansarse o sin aburrirse (como sí ocurriría con un ser humano). La computadora pudo adquirir mucha experiencia y se convirtió en el mejor jugador de damas de ese momento.¹⁰⁹

¹⁰⁸ Para efectos de esta investigación entendemos por nuevas tecnologías a: *Machine Learning*, *Deep Learning*, Big Data, Internet de las cosas y el cómputo en la nube.

¹⁰⁹ En el texto *Machine Learning: the power and promise of computers that learn by example* (The Royal Society, 2017) se explica que el uso de la IA para demostrar que las máquinas superaban a los hombres se dio desde esa época.

Samuel definió el ML como el campo de estudio que da a las computadoras la capacidad de aprender sin enseñarles explícitamente. En su artículo *Some Studies in Machine Learning Using the Game of Checkers* Samuel explica su experimento, el cual por su relevancia teórica se cita a continuación [traducción, síntesis y énfasis propio]:

Se han investigado con cierto detalle **dos procedimientos de aprendizaje automático** utilizando el juego de damas. Se ha trabajado lo suficiente para verificar el hecho de que una computadora puede ser programada para que aprenda a jugar un juego de damas **mejor que el que puede jugar la persona que escribió el programa**. Además, puede aprender a hacer esto en un período de **tiempo notablemente corto** (8 o 10 horas de tiempo de juego de la máquina) cuando se le dan solo las reglas del juego, un sentido de dirección y una lista redundante e incompleta de parámetros que se cree que tienen algo que ver con el juego, pero cuyos signos correctos y pesos relativos son desconocidos y no especificados. Los principios del aprendizaje automático verificados por estos experimentos son, por supuesto, aplicables a muchas otras situaciones... Los estudios presentados aquí se han referido a la programación de una computadora digital para que se comporte de una manera que, **si lo hicieran seres humanos o animales, se describiría como que involucra el proceso de aprendizaje... Tenemos a nuestro mando computadoras con la capacidad de manejo de datos adecuada y con suficiente velocidad computacional para hacer uso de técnicas de aprendizaje automático, pero nuestro conocimiento de los principios básicos de estas técnicas es aún rudimentario... Programar computadoras para aprender de la experiencia debería eventualmente eliminar la necesidad de gran parte de este esfuerzo de programación detallado...** [El método] Al principio, sería bueno distinguir claramente entre **dos enfoques generales del problema del aprendizaje automático**. Un método que podría llamarse el **Enfoque de la red neuronal** trata de la posibilidad de inducir el comportamiento aprendido en una red de conmutación conectada aleatoriamente (o su simulación en una computadora digital) como resultado de una rutina de recompensa y castigo. Un segundo enfoque, mucho más eficiente, es **producir el equivalente a una red altamente organizada que ha sido diseñada para aprender solo ciertas cosas específicas**. El primer método debería conducir al desarrollo de máquinas de aprendizaje de propósito general. Una comparación entre el tamaño de las redes de conmutación que pueden construirse o simularse razonablemente en la actualidad y el tamaño de las redes neuronales utilizadas por los animales sugiere que tenemos un largo camino por recorrer antes de obtener dispositivos prácticos. **El segundo procedimiento requiere reprogramación para cada nueva aplicación, pero es susceptible de realización en la actualidad. Los experimentos que se describirán aquí se basaron en este segundo enfoque.** [El problema] Durante algunos años [he] dedicado tiempo libre al tema del aprendizaje automático y [me he] concentrado en el desarrollo de procedimientos de aprendizaje aplicados a los juegos. **Un juego proporciona un vehículo conveniente para tal estudio en contraste con un problema tomado de la vida, ya que se eliminan muchas de las complicaciones de los detalles.** Se eligió el juego de damas, en lugar del ajedrez, porque la simplicidad de sus reglas permite poner mayor énfasis en las técnicas de aprendizaje... es justo afirmar que las damas contienen todas las características básicas de una actividad intelectual en la que los procedimientos heurísticos y los procesos de aprendizaje pueden desempeñar un papel importante y en la que estos procesos pueden ser evaluados.” ¹¹⁰

¹¹⁰ Cfr. Samuel, Arthur L., “Some Studies in Machine Learning Using the Game of Checkers”, en *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229, July 1959. Doi: 10.1147/rd.33.0210. Ver Anexo 1.

Samuel concluye que, como resultado de estos experimentos, se puede decir “...con cierta certeza que ahora es posible idear esquemas de aprendizaje que superarán en gran medida a una persona promedio y que dichos esquemas de aprendizaje pueden eventualmente ser económicamente viables si se aplican a problemas de la vida real.”¹¹¹

Antes de continuar con la exposición, es importante definir lo que es un modelo matemático y qué es un modelo en *Machine Learning* ya que los sistemas de IA se basan en ellos. Por modelo matemático, en su forma simple y para ilustrar nuestra investigación, es “una ecuación que describe cómo una o más variables de entrada se relacionan con una variable de salida. De esta forma, un modelo matemático es lo mismo que una función: un mapeo de entradas a salidas.”¹¹²

John D. Kelleher, hace una paráfrasis muy útil para seguir entendiendo este tema, nos dice: “George Box dice que todos los modelos son incorrectos, ¡pero algunos son útiles! Para que un modelo sea útil debe tener una correspondencia con el mundo real. Esta correspondencia es más obvia en términos del significado que se puede asociar con una variable. Por ejemplo, de forma aislada, un valor como 78,000 no tiene ningún significado porque no tiene correspondencia con los conceptos del mundo real. Pero el ingreso anual = \$ 78,000 nos dice cómo el número describe un aspecto del mundo real. Una vez que las variables en un modelo tienen un significado, podemos entender que el modelo describe un proceso a través del cual diferentes aspectos del mundo interactúan y provocan nuevos eventos. Los nuevos eventos son luego descritos por las salidas del modelo.”¹¹³

Una ecuación lineal “ $y = mx + c$ ” es un modelo simple, encontramos una variable de entrada (que es “x”) y una variable de salida (que es “y”); “m” y “c” son los dos parámetros del modelo que podemos establecer para ajustar la relación de datos de entrada y datos de salida. Pero ¿qué es un modelo en ML – DL?

¹¹¹ Ibidem., p. 548.

¹¹² John D., Kelleher, *Deep Learning, The MIT Press Essential Knowledge series*, Cambridge, The MIT Press, Massachusetts, 2019, p. 39-40.

¹¹³ Idem.

Un modelo de ML – DL, según Kelleher, es “un programa informático que codifica los patrones que el algoritmo de aprendizaje automático ha extraído de un conjunto de datos. Hay muchos tipos diferentes de modelos de aprendizaje automático; sin embargo, el aprendizaje profundo se centra en la creación de modelos de redes neuronales con múltiples capas de neuronas ocultas. Un modelo se crea (o entrena) ejecutando un algoritmo de aprendizaje automático en un conjunto de datos.”¹¹⁴

La segunda definición de ML, para efectos de esta investigación, la tomamos de Tom Mitchell, considerado como uno de los que perfeccionaron y divulgaron de este tema desde 1978, fecha en la que obtuvo su grado de doctor con una tesis relacionada con el aprendizaje automático. Mitchell conjuntamente con Ryszard Michalski y Jaime Carbonell construyeron los cimientos conceptuales para el desarrollo acelerado del ML.¹¹⁵

Mitchell comenta: ¹¹⁶: “...we say that a machine learns with respect to a particular task T, performance metric P, and type of experience E, if the system

¹¹⁴ Cfr. John D., Kelleher, op. cit., 253. Kelleher complementa su comentario con lo siguiente: “Una vez que se ha entrenado el modelo, se puede utilizar para analizar nuevas instancias; el término inferencia a veces se usa para describir el proceso de analizar una nueva instancia usando un modelo entrenado. En el contexto del aprendizaje automático, los términos modelo y función a menudo se usan indistintamente: un modelo es una instanciación de una función como un programa de computadora.” [La traducción es propia]

¹¹⁵ Cfr. Ryszard Michalski, Jaime Carbonell y T. Mitchell, *Machine Learning. an Artificial Intelligence approach*, Tioga Publishing Co., Volume 2, Palo Alto, 1983. https://www.researchgate.net/publication/238582946_Machine_learning_an_Artificial_Intelligence_approach_Volume_2 [Considerado el primer texto que genera una propuesta muy fundamentada para el uso del ML. El segundo se publicó por Mitchell en 1997]

¹¹⁶ Cfr. Tom, Mitchell, *Machine Learning*, Mc Graw Hill Science, s/l, 1997, p. 2 y ss. Tom Mitchell es profesor en el Departamento de *Machine Learning* en la Carnegie Mellon University es considerado uno de los pioneros en el tema de ML. (<https://www.lti.cs.cmu.edu/>). El departamento tiene el *Language Technologies Institute* que busca realizar investigaciones en las áreas de procesamiento del lenguaje natural, lingüística computacional, extracción de información, resumen y respuesta a preguntas, recuperación de información, minería de textos y análisis, representación del conocimiento, razonamiento y adquisición de tecnologías del lenguaje para la educación, el aprendizaje automático, la traducción automática, la computación e interacción multimodal, el procesamiento del habla y las interfaces habladas y el procesamiento del diálogo. Mitchell tiene 2 líneas de investigación basadas en cómo construir computadoras que aprendan de la experiencia (cómo formular automáticamente hipótesis para la resolución de problemas). Sus dos proyectos son: Enfoques de aprendizaje automático para analizar la actividad del cerebro humano, utilizando imágenes de resonancia magnética funcionales (MRI) para capturar imágenes tridimensionales de la actividad del cerebro humano en donde existe una gran necesidad de creación de nuevos algoritmos para analizar datos. Este proyecto plantea interesantes preguntas de aprendizaje

reliably improves its performance P at task T , following experience E ".¹¹⁷ Mitchell definió el aprendizaje automático poniendo un ejemplo de un programa de cómputo: "...un programa de computadora aprende de la experiencia (experience = " E "), con respecto a alguna tarea (Task = " T ") y alguna medida del desempeño (performance = " P "), si su desempeño en T , medido por P , mejora con la experiencia E ." Ejemplo: El juego de damas.

E = la experiencia de jugar muchos juegos de damas

T = la tarea de jugar damas.

P = la probabilidad de que el programa gane el próximo juego.

En el experimento de damas hecho por Samuel, la experiencia E sería la experiencia del programa jugando decenas de miles de partidas contra sí mismo. La tarea T sería la tarea de jugar damas, y la medida del desempeño P será la probabilidad de ganar la próxima partida contra algún oponente nuevo.

Mitchell pone otro ejemplo relacionado con el detector de spam del correo electrónico (sabe qué debe permitir llegar a la bandeja de entrada y qué detener y mover a la bandeja de correo no deseado por ser spam). En un e-mail se puede seleccionar algún mensaje que no corresponde como spam o desmarcar otro que sí corresponde es spam. Y basándose en los correos marcados como spam, el programa aprende a filtrar otros correos spam ¿Cuál es la Tarea T aquí? La respuesta es: clasificar los correos de spam y no spam.¹¹⁸

automático, por ejemplo, cómo entrenar clasificadores en datos extremadamente ruidosos y de dimensiones elevadas, y cómo aprender modelos temporales que caracterizan la evolución de estados cognitivos ocultos mientras los humanos realizan tareas como leer y responder preguntas. La segunda línea de investigación es sobre Asistentes de estaciones de trabajo inteligentes que aprenden a ayudar a sus usuarios, donde se busca encontrar un agente de software que pueda comprender el correo electrónico, el calendario, los archivos de texto y las acciones del usuario, y que pueda conocer los intereses, hábitos y tareas del usuario para ayudarlo en una amplia gama de actividades de la vida cotidiana. Consúltese: <https://www.lti.cs.cmu.edu/people/15355/tom-mitchell> [20 de julio de 2022]

¹¹⁷ Cfr. Tom Mitchell, *The Discipline of Machine Learning*, School of Computer Science Carnegie Mellon University, Pittsburgh, 2016. <http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf> [20 de julio de 2022]

¹¹⁸ *Idem*. Para conocer más detalles de estos temas, existe un curso de Mitchell que es público en: <https://scs.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=257476ca-e66a-4cdb-9ffc-8abdd3129954> [20 de julio de 2022]

En palabras de Tom Mitchell: “Un campo científico se define por la cuestión central que estudia. El campo del *Machine Learning* busca dar una respuesta a una pregunta: “¿Cómo podemos construir sistemas informáticos que mejoren automáticamente con la experiencia y cuáles son las leyes fundamentales que gobiernan todos los procesos de aprendizaje?”. Esta pregunta cubre una amplia gama de tareas de aprendizaje; [por ejemplo] diseñar robots móviles autónomos que aprendan a navegar a partir de su propia experiencia, extraer datos de registros médicos históricos para saber qué pacientes futuros responderán mejor a qué tratamientos y construir motores de búsqueda que se personalizan automáticamente según los intereses de sus usuarios.

Para ser más precisos, decimos que una máquina aprende con respecto a una tarea particular *T*, métrica de desempeño *P* y tipo de experiencia *E*, si el sistema mejora confiablemente su desempeño *P* en la tarea *T*, siguiendo la experiencia *E*. Dependiendo de cómo especifiquemos *T*, *P* y *E*, la tarea de aprendizaje también se puede llamar por nombres como minería de datos, descubrimiento autónomo, actualización de bases de datos, programación, etc. Mitchell expuso desde 1997, que el campo del ML se ocupa de cuestiones sobre cómo construir programas informáticos que mejoren por sí mismos -y con la experiencia cotidiana de su operación- las actividades que se les encomienden. Para ello, se han producido importantes avances en la teoría del ML y los algoritmos que usa. En síntesis:

“Machine Learning addresses the question of how to build computer programs that improve their performance at some task through experience....”

- *Machine Learning* algorithms have proven to be of great practical value in a variety of application domains. They are especially useful in (a) data mining problems where large databases may contain valuable implicit regularities that can be discovered automatically (e.g., to analyze outcomes of medical treatments from patient databases or to learn general rules for credit worthiness from financial databases); (b) poorly understood domains where humans might not have the knowledge needed to develop effective algorithms (e.g., human face recognition from images); and (c) domains where the program must dynamically adapt to changing conditions (e.g., controlling manufacturing processes under changing supply stocks or adapting to the changing reading interests of individuals).

- *Machine Learning* draws on ideas from a diverse set of disciplines, including artificial intelligence, probability and statistics, computational complexity, information theory, psychology and neurobiology, control theory, and philosophy.

- A well-defined learning problem requires a well-specified task, performance metric, and source of training experience.

- Designing a *Machine Learning* approach involves a number of design choices, including choosing the type of training experience, the target function to be learned, a representation for this target function, and an algorithm for learning the target function from training examples.

- Learning involves search: searching through a space of possible hypotheses to find the hypothesis that best fits the available training examples and other prior constraints or knowledge. Much of this book is organized around- different learning methods that search different hypothesis spaces (e.g., spaces containing numerical functions, neural networks, decision trees, symbolic rules) and around theoretical results that characterize conditions under which these search methods converge toward an optimal hypothesis.”¹¹⁹

La teoría del aprendizaje automático intenta responder preguntas como ¿Cómo varía el rendimiento del aprendizaje con la cantidad de ejemplos de capacitación presentados? y ¿Qué algoritmos de aprendizaje son los más apropiados para varios tipos de tareas de aprendizaje?¹²⁰

¹¹⁹ “El aprendizaje automático aborda la cuestión de cómo *crear programas de computadora que mejoren su desempeño* en alguna tarea a través de la experiencia...:

- Los algoritmos de aprendizaje automático han demostrado ser de gran valor práctico en una variedad de dominios de aplicaciones. Son especialmente útiles en (a) problemas de minería de datos donde grandes bases de datos pueden contener valiosas regularidades implícitas que se pueden descubrir automáticamente (por ejemplo, para analizar los resultados de tratamientos médicos de bases de datos de pacientes o para aprender reglas generales de solvencia crediticia de bases de datos financieras); (b) dominios poco entendidos donde los humanos podrían no tener el conocimiento necesario para desarrollar algoritmos efectivos (por ejemplo, reconocimiento de rostros humanos a partir de imágenes); y (c) dominios en los que el programa debe adaptarse dinámicamente a las condiciones cambiantes (por ejemplo, controlando los procesos de fabricación cuando las existencias de suministro cambiantes o adaptándose a los intereses de lectura cambiantes de los individuos).

- El aprendizaje automático se basa en ideas de un conjunto diverso de disciplinas, que incluyen inteligencia artificial, probabilidad y estadística, complejidad computacional, teoría de la información, psicología y neurobiología, teoría del control y Filosofía.

- Un problema de aprendizaje bien definido requiere una tarea, una métrica de desempeño y una fuente de experiencia de capacitación bien especificadas.

- El diseño de un enfoque de aprendizaje automático implica una serie de opciones de diseño, incluida la elección del tipo de experiencia de entrenamiento, la función objetivo que se aprenderá, una representación de esta función objetivo y un algoritmo para aprender la función objetivo a partir de ejemplos de entrenamiento.

- El aprendizaje implica búsqueda: buscar a través de un espacio de posibles hipótesis para encontrar la hipótesis que mejor se ajusta a los ejemplos de formación disponibles y otras limitaciones o conocimientos previos. Gran parte de este libro está organizado en torno a diferentes métodos de aprendizaje que buscan diferentes espacios de hipótesis (por ejemplo, espacios que contienen funciones numéricas, redes neuronales, árboles de decisión, reglas simbólicas) y en torno a resultados teóricos que caracterizan las condiciones bajo las cuales estos métodos de búsqueda convergen hacia una hipótesis óptima.” Cfr. Mitchell, Tom M. (1997). *Machine Learning*. s/l. McGraw-Hill Science pp. 17-18.

¹²⁰ Mitchell, Tom M. *Machine Learning*, op. cit., Introduction.

Otra definición de ML construida a petición de la Comisión Europea la encontramos en el *Libro Blanco sobre la Inteligencia Artificial* el cual dice lo siguiente: “constituye un subapartado de la IA, los algoritmos son entrenados para inferir determinados modelos a partir de un conjunto de datos, a fin de determinar las acciones que se requieren para alcanzar un objetivo determinado. Los algoritmos pueden seguir aprendiendo mientras se utilizan.”¹²¹

Y finalmente, visto este tema desde la sociedad civil, organizaciones como Amnistía Internacional y Access Now la definen en la Declaración de Toronto al momento de explicar el aprendizaje automático o *Machine Learning* como un subconjunto de la inteligencia artificial que tiene “la capacidad de aprender y mejorar de forma automática a partir de las experiencias, sin necesidad de programación explícita”.¹²²

Definiciones sobre “*Machine Learning*” o “Aprendizaje Automático”

Autor	Definición
Artur Samuel	...el campo de estudio que da a las computadoras la capacidad de aprender sin enseñarles explícitamente...
Tom Mitchell	...un programa de computadora aprende de la experiencia (experience = “E”), con respecto a alguna tarea (Task = “T”) y alguna medida del desempeño (performance = “P”), si su desempeño en T, medido por P, mejora con la experiencia E.
Declaración de Toronto	...la capacidad de aprender y mejorar de forma automática a partir de las experiencias, sin necesidad de programación explícita...
Libro blanco	En lo que se refiere a las técnicas de aprendizaje automático, que constituyen un subapartado de la IA, los algoritmos son entrenados para inferir determinados modelos a partir de un conjunto de datos, a fin de determinar las acciones que se requieren para alcanzar un objetivo determinado. Los algoritmos pueden seguir aprendiendo mientras se utilizan.

Fuente: Elaboración propia, con base en las definiciones expuestas previamente.

Uno de los riesgos más importantes que presenta el aprendizaje automático es el peligro de amplificar la discriminación y los sesgos existentes en ciertos grupos vulnerables de la población humana. Cuando se utilizan datos históricos para

¹²¹ Comisión Europea, *Libro Blanco sobre la inteligencia artificial...*, op. cit., p. 20

¹²² The Toronto Declaration: Protecting the rights to equality and non-discrimination in *Machine Learning* systems: <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/> [20 de julio de 2022]

entrenar a los sistemas de aprendizaje automático, estos sistemas pueden reafirmar o incluso aumentar el sesgo estructural existente. También se pueden causar daños discriminatorios cuando las decisiones tomadas al diseñar los sistemas de IA producen resultados sesgados, ya sea de forma deliberada o no.

¿Cómo funciona el ML – DL? Para efectos de diferenciar los detalles de cada uno, haremos una separación conceptual. Primero explicaremos el ML y posteriormente el DL. No obstante, para efectos del trabajo los consideramos como parte de esa IA que no es controlada por seres humanos y que afecta sus derechos.

Los métodos de aprendizaje automático son útiles en entornos de aplicaciones donde las personas no pueden proporcionar especificaciones precisas para el comportamiento deseado del programa. Tales especificaciones precisas las podemos apreciar, por ejemplo, en el reconocimiento óptico de caracteres, en el reconocimiento de escritura a mano, en el reconocimiento de voz, la dirección automatizada de automóviles y el control y la navegación de robots. Una propiedad clave de estas tareas es que las personas pueden realizarlas con bastante facilidad, pero no pueden articular exactamente cómo las realizan. Por lo tanto, las personas pueden proporcionar las variables de entrada y salida, pero no pueden proporcionar especificaciones o algoritmos precisos.

Los métodos de aprendizaje automático también son apropiados para situaciones en las que la tarea cambia con el tiempo o entre diferentes usuarios, de modo que un programador no puede anticipar exactamente cómo debería comportarse el programa. Por ejemplo, se han utilizado métodos de aprendizaje automático para evaluar el riesgo de las tarjetas de crédito, filtrar artículos de noticias, refinar las consultas de recuperación de información y predecir el comportamiento de navegación en la World Wide Web.

Una tercera área de aplicación de los algoritmos de aprendizaje automático es el problema de encontrar patrones importantes para una actividad específica en las bases de datos, a veces llamado minería de datos. Muchas corporaciones recopilan información sobre las compras de los clientes, las reclamaciones

presentadas por los proveedores médicos, las reclamaciones de seguros presentadas por los conductores, los registros de mantenimiento de aeronaves, etc. Los algoritmos de aprendizaje automático (y también muchos métodos estadísticos tradicionales) pueden encontrar patrones importantes en estos datos que pueden aplicarse para mejorar el marketing, detectar fraudes y predecir problemas futuros.

Otra característica es que *Machine Learning* se basa en definir reglas de comportamiento mediante el examen y la comparación de grandes conjuntos de datos para encontrar patrones comunes. Por ejemplo, podemos proporcionar un programa de aprendizaje automático con un gran volumen de fotos de algún animal y entrenar al modelo para que regrese la etiqueta “gato” cuando ha proporcionado una imagen de un gato. También se puede crear una etiqueta para perro y proporcionar fotos de perros para entrenarlo, entre más muestras se proporcione al algoritmo de aprendizaje supervisado más preciso se vuelve al clasificar los nuevos datos (es un ejemplo de AAS que se verá más adelante).

Hay varios tipos de algoritmos de ML, los principales son los que la doctrina ha clasificado en tres categorías: 1) aprendizaje supervisado, 2) aprendizaje no supervisado y 3) aprendizaje por refuerzo¹²³.

En el aprendizaje supervisado, un sistema se entrena con datos etiquetados. Las etiquetas clasifican cada punto de datos en uno o más grupos, como peras o manzanas. El sistema aprende cómo se estructuran estos datos, conocidos como datos de entrenamiento, y los utiliza para predecir las categorías de datos nuevos o de prueba. El aprendizaje no supervisado es aprender sin etiquetas; es decir, su objetivo es detectar las características que hacen que los puntos de datos sean más o menos similares entre sí, por ejemplo, creando grupos y asignando datos a estos grupos. Y, el aprendizaje por refuerzo se centra en aprender de la experiencia y se encuentra entre el aprendizaje supervisado y el no supervisado. En un entorno típico de aprendizaje por refuerzo, un agente interactúa con su entorno y recibe una

¹²³ Thomas G. Dietterich, “*Machine Learning*”, *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., Londres, 2013.

función de recompensa que intenta optimizar, por ejemplo, el sistema podría ser recompensado por ganar un juego.

Adicionalmente, cuando se implementan sistemas de aprendizaje automático, existe una distinción clave entre aquellos que se hacen fuera de línea (*offline*) y en línea (*online*). Los sistemas de aprendizaje sin conexión (*offline*) se entrenan y prueban en un entorno fuera de red, y los modelos entrenados se “congelan” antes de implementarlos en un entorno en línea. Cualquier capacitación posterior también se realizará en un entorno fuera de línea, se probará y luego se implementará utilizando métodos convencionales de administración de cambios de software. Este enfoque es el más común en la actualidad, ya que brinda una oportunidad para la verificación humana del sistema, antes de que el sistema interactúe con cualquier usuario. Existen también los sistemas de aprendizaje en línea (*online*) también se entrenan y prueban en un entorno fuera de línea antes de la implementación, pero los algoritmos de aprendizaje continúan aplicándose al modelo entrenado después de la implementación. Esto significa que el desempeño del sistema “en la naturaleza” puede seguir mejorando en tiempo real en respuesta a los datos del mundo real. También significa que no existe la posibilidad de que las personas comprueben las consecuencias de las actualizaciones del modelo antes de que los usuarios estén expuestos a ellas. En general, cualquier problema de aprendizaje automático se puede asignar a una de las técnicas: aprendizaje supervisado, aprendizaje no supervisado y por refuerzo.¹²⁴

A continuación, explicaremos con mayor detalle cada uno de los tres tipos de aprendizaje automático (supervisado, no supervisado y por refuerzo).

¿Qué es el Aprendizaje Supervisado (AS)? El término AS tiene que ver con dar a la máquina un conjunto de datos conocidos para que los procese el algoritmo, los cuales son las “respuestas correctas”. El AS se dividen a su vez en dos tipos: problemas de regresión, problemas de clasificación. Por ejemplo, la venta de un

¹²⁴ Cfr. The Royal Society, *Machine Learning: the power and promise of computers that learn by example*, The Royal Society, Londres, 2017. Consúltase en: <https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

departamento. Aquí primero se calcula el precio del departamento tomando en cuenta el valor de otros departamentos que existen en determinada zona con características similares para ubicar si el precio de venta es el más adecuado. El usuario o el programador indica el precio de mercado de otros departamentos y la tarea del algoritmo es buscar más información de esas “respuestas correctas” para así ofertar el departamento que se quiere vender a un precio específico.

Este ejemplo es conocido como un problema de regresión, es decir, cuando intentamos predecir una salida de valor continuo (en el ejemplo es el precio que tienen otros departamentos). El término regresión se refiere al hecho de intentar predecir un atributo de valores continuos.

Un segundo ejemplo de AS lo podemos ver en registros médicos para tratar de predecir si un tumor es benigno o maligno. Si el médico descubre un tumor y lo procesa con AS el sistema que se utilice tiene ya programado un conjunto de datos sobre ese tema. Entonces esto se puede graficar de la siguiente manera:

Tumor benigno = 0

Tumor maligno = 1

Eje horizontal = Tamaño del tumor

Eje vertical = un uno y un cero (sí o no), para los tumores malignos o benignos.

Un tumor de “x” tamaño será benigno, uno de “y” tamaño será maligno y así sucesivamente, (según la información que previamente se ingrese a la máquina). Pensemos que una persona tiene un tumor y que el tamaño del tumor está en algún punto de estos valores. La pregunta de AS es ¿cuál es la probabilidad de que un tumor sea maligno o benigno?

Este es un ejemplo de un problema de clasificación. El término clasificación se refiere al hecho de que intentamos predecir un valor discreto, una salida cero o uno, maligno o benigno. En problemas de clasificación, algunas veces puedes tener más de dos posibles valores para la salida. Como ejemplo concreto, puede haber tres tipos de cáncer (o más). De esta forma se puede intentar predecir

un valor discreto de cero, uno, dos o tres; donde cero indica benigno o sin cáncer; uno y dos pueden significar distintos tipos de cáncer (o gravedad del cáncer). Esto también es considerado como un problema de clasificación por el conjunto de valores discretos de salida correspondiente (sin cáncer o tipo uno, con cáncer tipo dos, o con cáncer tipo tres. En resumen, los algoritmos de AS nos dicen cuál es la respuesta correcta. El precio de un departamento, saber si un tumor es maligno o benigno, etcétera.

En cuanto al problema de regresión, podemos decir que el objetivo es la predicción de una salida de valor continuo. Supongamos que una empresa desea desarrollar algoritmos de aprendizaje para abordar estos problemas. En el problema de regresión, tienes un gran inventario de artículos idénticos para vender, y quieres predecir cuántos artículos vas a vender en los tres meses próximos. En el segundo problema, de regresión, tienes muchos usuarios, y deseas construir un software para examinar cada una de las cuentas de tus clientes. Decidir si la cuenta ha sido hackeada o no tiene problemas.

El número de artículos que vendo es un valor continuo. En cambio, en el segundo caso, se trata de un problema de clasificación, porque podría decir que pongo el valor que quiero predecir como cero para aquellas cuentas que no han sido hackeadas y el valor uno para aquellas cuentas que sí han sido hackeadas. Al igual que los cánceres de mama donde cero es benigno, y uno es maligno.

Entonces, debe quedar claro que, en el aprendizaje supervisado, se nos proporciona un conjunto de datos y ya sabemos cómo debería ser nuestra salida o respuesta correctas; se tiene idea de la entrada y la salida. Los problemas de aprendizaje supervisados se clasifican en problemas de “regresión” y “clasificación”. En un problema de regresión, estamos tratando de predecir resultados dentro de una salida continua, lo que significa que estamos tratando de asignar variables de entrada a alguna función continua. En cambio, en un problema de clasificación,

estamos tratando de predecir resultados en una salida discreta. En otras palabras, estamos tratando de asignar variables de entrada en categorías discretas¹²⁵.

Ejemplo 1:

Dada la información sobre el tamaño de los departamentos en una determinada ciudad, se intentará predecir su precio. El precio en función del tamaño es una salida continua, por lo que este es un problema de regresión. Se puede convertir este ejemplo en un problema de clasificación al hacer nuestra predicción sobre si el departamento se vende por más o menos que el precio de venta. Aquí estamos clasificando las casas en función del precio en dos categorías discretas.

Ejemplo 2:

Regresión: dada una imagen de una persona, tenemos que predecir su edad

Clasificación: dado un paciente con un tumor, tenemos que predecir si el tumor es maligno o benigno.

En resumen, el proceso de AS se desarrolla de la siguiente manera:

1. Se define el tema (Venta de departamentos)
2. Se ubica el conjunto de datos (cantidad de departamentos de diferentes tamaños y que se vendieron en un rango de precios diferentes).
3. Se ubican las características del tema (un departamento, para venta, con determinado número de metros cuadrados, determinada zona geográfica).
4. La información se da al algoritmo para que haga la búsqueda (buscar el precio para la venta). Aquí interviene el algoritmo de “aprendizaje supervisado”.

En el AS hay una “respuesta correcta” para cada cálculo. Es decir, los datos contienen el precio real de cada casa y el número de metros más el número de

¹²⁵ Dietterich comenta: *In supervised learning, the goal is to learn the form of a function $y = f(x)$ by analyzing examples of the form $(x; y_i)$, where $Y_i = f(x_i)$. Each input value X is usually an n -dimensional vector, where each dimension is either discrete or real-valued.*

metros que tiene la casa que se quiere vender. Se predice el resultado a través del valor real.

¿Qué es el Aprendizaje No Supervisado (ANS)? En cada ejemplo de Aprendizaje Supervisado, teníamos información considerada explícitamente como una “respuesta correcta” (Tumor benigno o maligno, por ejemplo). En el ANS solamente hay datos. Estos datos no tienen etiquetas específicas o no están agrupados con la misma etiqueta. Pensemos que vamos a una empresa a digitalizar la información y que esa información se encuentra desorganizada, dispersa, sin etiquetas, sin un orden. Solo son datos, una gran cantidad de datos que la empresa desea digitalizar y utilizar para sus actividades comerciales. La empresa no nos dice qué hacer con ellos y no nos dicen qué es cada dato puntual. En lugar de eso, nos dicen este es el conjunto de datos, ustedes encárguense de ordenarlos y de decirnos qué podemos hacer con ellos.

Un algoritmo de ANS, con esos datos, podría decidir que los datos están en dos o más grupos (clasificarlos por categorías). Y si el algoritmo puede dividir estos datos en dos o más grupos se le conoce como “Algoritmo de Agrupamiento”. Un ejemplo muy conocido son los buscadores de noticias como “Google New México”. Si uno quisiera revisar los avances del COVID-19, lo que Google News hace es analizar miles o cientos de miles de noticias en la web y agruparlas en noticias relacionadas con temas específicos.

Otro ejemplo lo encontramos en el área médica y está relacionado con entender el ADN. El ADN tiene datos llamados técnicamente “micro matriciales” ¹²⁶. En algunos experimentos se junta un grupo de personas diferentes para medir qué cantidad de genes tienen y si hay alguna mutación. Aquí se puede utilizar un algoritmo de agrupamiento para su análisis, es decir, se utiliza el ANS porque no le decimos al algoritmo (previamente) que las personas son de tal o cual tipo.

¹²⁶ La micro matriz de ADN es una herramienta para determinar si el ADN de una persona contiene una mutación genética. Cfr. Tecnología de micro matriz de ADN <https://www.genome.gov/es/about-genomics/fact-sheets/Tecnologia-de-micromatriz-de-ADN> [15 agosto de 2022]

El ANS se usa para organizar grandes agrupamientos de datos que pueden estar en un solo equipo o en varios equipos informáticos o en varios equipos de varias organizaciones o con datos de redes sociales para hacer que el trabajo sea más eficiente. Esta segunda aplicación es sobre el análisis de redes sociales.

Muchas empresas tienen enormes bases de datos con información de sus clientes y buscan descubrir segmentos de mercado, agrupar a clientes en diferentes segmentos de ese mercado, para vender o comercializar de forma automática y más eficiente los diferentes segmentos. Se tienen los datos de clientes, pero no se sabe en cuáles y cuántos segmentos de mercado. El uso de algoritmos del ANS para descubrir todo esto a partir de los datos es de gran utilidad.¹²⁷

La abundancia de datos permite un aprendizaje no supervisado potente para campos que pueden servir para reconocer patrones, hacer minería de datos, crear perfiles, descubrir conocimiento, hacer análisis predictivo.

El tema de grandes agrupamientos de datos está relacionado con el uso de patrones oscuros que se pretende regular para prohibir que las grandes plataformas como Facebook las utilicen. El término patrones oscuros se utiliza para describir interfaces en línea en sitios web y aplicaciones diseñadas para manipular intencionalmente a los usuarios para que tomen decisiones que de otro modo no tomarían en circunstancias normales. Estas tácticas de diseño, extraídas de una extensa investigación de psicología conductual, son utilizadas con frecuencia por

¹²⁷ Dietterich comenta: Unsupervised learning algorithms attempt to find structure in a set of data points. Consider, for example, an astronomy project in which the spectra of thousands of stars have been gathered. Each spectrum is a vector of numbers giving the intensity of a star's light at various wavelengths. Can we find patterns in this data? Suppose, for example, that we believe that there are different classes of stars. Could we use this data to identify the different classes of stars and group the stars into those classes? Unsupervised learning often involves clustering data into classes in this fashion. [Los algoritmos de aprendizaje no supervisados intentan encontrar estructura en un conjunto de puntos de datos. Considere, por ejemplo, un proyecto de astronomía en el que se han reunido los espectros de miles de estrellas. Cada espectro es un vector de números que da la intensidad de la luz de una estrella en varias longitudes de onda. ¿Podemos encontrar patrones en estos datos? Supongamos, por ejemplo, que creemos que existen diferentes clases de estrellas. ¿Podríamos usar estos datos para identificar las diferentes clases de estrellas y agrupar las estrellas en esas clases? El aprendizaje no supervisado a menudo implica agrupar datos en clases de esta manera.] Traducción propia.

las plataformas de redes sociales para guiar (algunos dicen engañar) a los consumidores para que acepten entornos o prácticas de determinadas empresas.¹²⁸

¿Qué es el Aprendizaje por Refuerzo (AR)?¹²⁹ El tercer tipo de algoritmo de aprendizaje automático por refuerzo, el cual se basa en proporcionar al algoritmo con un conjunto de reglas y restricciones, y dejarle aprender cómo lograr sus objetivos (ensayo-error). Aquí se define lo deseado, las acciones permitidas y las restricciones. El algoritmo descubre cómo lograr el fin deseado probando diferentes combinaciones de acciones permitidas, y es recompensado o castigado dependiendo de si la decisión fue buena. El algoritmo hace todo lo posible para maximizar sus recompensas dentro de las limitaciones proporcionadas. AlphaGo de Google (DeepMind), trabaja mediante aprendizaje por refuerzo.

Los problemas de aprendizaje reforzado surgen, según Dietterich (2003), cuando un programa de computadora debe tomar una secuencia de decisiones antes de recibir una recompensa al final. Pensemos en una computadora que está compitiendo en un juego de damas o de ajedrez. La computadora debe programarse para realizar una larga secuencia de movimientos para saber si gana o pierde el juego y logra obtener su recompensa.

¹²⁸ Los senadores Warner (D-VA) y Deb Fischer (R-NE), presentaron en 2019 una iniciativa de Ley de Reducción de Usuarios de Experiencias Engañosas para Usuarios en Línea (DETOUR), con el fin de transparentar este mercado y garantizar que los consumidores puedan tomar decisiones más informadas sobre cómo y cuándo compartir su información personal. “Cualquier política de privacidad que implique consentimiento se debilita por la presencia de patrones oscuros. Estas interfaces de usuario manipuladoras limitan intencionalmente la comprensión y socavan la elección del consumidor. Las indicaciones engañosas para hacer clic en el botón “Aceptar” a menudo pueden transferir sus contactos, mensajes, actividad de navegación, fotos o información de ubicación sin que se dé cuenta. Nuestra legislación busca frenar el uso de estas interfaces deshonestas y aumentar la confianza en línea”, dijo el senador Fischer, miembro del Comité de Comercio del Senado... Los patrones oscuros... confunden o atrapan a los usuarios para compartir información personal excesivamente o impulsar el uso compulsivo, especialmente de los usuarios más vulnerables, incluidos los niños”, dijo Tristan Harris, cofundador del Centro de Tecnología Humana”. Cfr. Tristan Harris *Potential Policy Proposals for Regulation of Social Media and Technology Firms* Consúltense más detalles en: <https://www.warner.senate.gov/public/cache/files/d/3/d32c2f17-cc76-4e11-8aa9-897eb3c90d16/65A7C5D983F899DAAE5AA21F57BAD944.social-media-regulation-proposals.pdf> [20 de julio de 2022]

¹²⁹ Hernández Sánchez, José Luis, *Apuntes del Curso Introducción a la Inteligencia Artificial* (Impartido por: Rav Ahuja, *AI and Data Science Program Director IBM*) s/e, México, 2020.

Los algoritmos de aprendizaje reforzado están diseñados para resolver este tipo de aprendizaje y aprenden una tarea al realizarla. Antes de cada movimiento, el algoritmo examina el estado actual del mundo (es decir, la posición actual del tablero en un juego de ajedrez) y luego elige una acción a (es decir, un movimiento de ajedrez). La acción hace que el mundo cambie a un nuevo estado. Después de cada transición, la computadora recibe una recompensa. En el ajedrez, la recompensa es cero hasta el final del juego, donde es 1 (ganar), 0 (empatar) o -1 (perder).¹³⁰

El éxito del ML se debe a la posibilidad que da de explotar los datos de la organización (que genera y que por lo regular se encuentran estructurados) y los datos que están disponibles en fuentes abiertas, como imágenes, voz, video, documentos, correos electrónicos, redes sociales, blogs, representan una fuente significativa de oportunidad para las empresas y su relación con clientes, usuarios, competidores, productos, mercados, diagnósticos, etcétera. Estos datos proporcionan a los sistemas de aprendizaje automático ingredientes que de otra manera el ser humano no podría obtener para mejorar sus actividades.¹³¹

B) Deep Learning

¿Qué es *Deep Learning* (DL) o la revolución de la IA basada en redes neuronales convolucionales profundas? Dentro de los algoritmos clasificados “por refuerzo” existe lo que llama: Técnica de DL o algoritmos de DL basados en Redes Neuronales Convolucionales Profundas. El DL es “un subcampo del aprendizaje automático que diseña y evalúa algoritmos de entrenamiento y arquitecturas para modelos de redes neuronales *profundas modernas*”¹³². Resaltamos aquí a dos conceptos “profundas” y “modernas” ya que al momento de su creación la red neuronal artificial funcionaba con capas simples (no con múltiples capas). En

¹³⁰ Dietterich, Thomas G, “*Machine Learning*”, *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., Londres, 2013, pp. 1056–1059. <https://dl.acm.org/doi/pdf/10.5555/1074100.1074563> [20 de julio de 2022]

¹³¹ The Royal Society, *Machine Learning...*, *op. cit.*, Executive summary, pp. 5 y 6.

¹³² Cfr. John D., Kelleher, *Deep Learning*, *The MIT Press Essential Knowledge series*, Cambridge, The MIT Press, Massachusetts, 2019, p. 252.

cambio, una Red Neuronal Profunda (RNP) “es una red de neuronas que tiene *múltiples capas*”.¹³³

En 1958, se crea el primer algoritmo basado en la unidad primigenia de nuestro cerebro: la neurona. Su creador fue Frank Rosenblatt y lo denominó “Perceptrón”. Su estudio fue basado en cuatro artículos científicos clave de Ashby, Hayek, Hebb y Uttley¹³⁴, que se publicaron entre 1949 y 1956, y que buscaban hacer que las máquinas funcionaran conforme lo hace el cerebro humano. Además, entendía que desafortunadamente, el lenguaje de la lógica simbólica y el álgebra booleana no era el camino adecuado para hacer investigaciones de este tipo. Por su relevancia transcribimos el texto:

“...Si finalmente queremos comprender la capacidad de los organismos superiores para el reconocimiento perceptivo, la generalización, el recuerdo y el pensamiento, primero debemos tener respuestas a tres preguntas fundamentales:

1. ¿Cómo percibe o detecta el sistema biológico la información sobre el mundo físico?
2. ¿De qué forma se almacena o recuerda la información?
3. ¿Cómo influye la información contenida en el almacenamiento o en la memoria en el reconocimiento y el comportamiento?

La primera de estas preguntas pertenece al campo de la fisiología sensorial y es la única para la que se ha logrado una comprensión apreciable. Este artículo se ocupará principalmente de la segunda y tercera preguntas, que todavía están sujetas a una gran cantidad de especulaciones, y donde los pocos hechos relevantes que proporciona actualmente la neurofisiología aún no se han integrado en una teoría aceptable. Con respecto a la segunda pregunta, se han mantenido dos posiciones alternativas. La primera sugiere que el almacenamiento de información sensorial se realiza en forma de representaciones o imágenes codificadas, con algún tipo de mapeo uno a uno entre el estímulo sensorial y el patrón almacenado. [...] Esta hipótesis es atractiva por su simplicidad y fácil inteligibilidad, y se ha desarrollado una gran familia de modelos cerebrales teóricos en torno a la idea de una memoria representativa codificada. El enfoque alternativo, que se deriva de la tradición del empirismo británico, pone en peligro la suposición de que las imágenes de los estímulos tal vez nunca se registren en absoluto, y que el sistema nervioso central simplemente actúa como una intrincada red de conmutación, donde la retención toma la forma de nuevos cambios, conexiones, o vías, entre centros de actividad. En muchos de los desarrollos más recientes de esta posición [...] las “respuestas” que están asociadas a los estímulos pueden estar completamente contenidas dentro del propio SNC. En este caso la respuesta representa una “idea” más que una acción. La característica importante de este enfoque es que nunca hay un simple mapeo del estímulo en la memoria, de acuerdo con algún código que permita su posterior reconstrucción. Cualquier información que se retenga debe almacenarse de alguna manera como preferencia para una respuesta particular, es decir, la información está

¹³³ *Idem.*

¹³⁴ W. R., Ashby, *Design for a brain*, Wiley, New York, 1952. F. A., Hayek, *The sensory order*, Chicago Press, Chicago, 1952. D. O. Hebb, *The organization of behavior*. Wiley, New York, 1949. A. M., Uttley, “Conditional probability machines and conditioned reflexes”, en C. E. Shannon & J. McCarthy (Eds.), *Automata studies*. Princeton University Press, Princeton, 1956, pp.253-275.

contenida en conexiones o asociaciones en lugar de representaciones topográficas. (El término respuesta, para el resto de esta presentación, debe entenderse como cualquier estado distinguible del organismo, que puede involucrar o no una actividad muscular detectable externamente. La activación de algún núcleo de células en el sistema nervioso central, por ejemplo, puede constituir una respuesta, según esta definición).

Correspondiendo a estas dos posiciones sobre el método de retención de la información, existen dos hipótesis con respecto a la tercera pregunta, la forma en que la información almacenada ejerce su influencia sobre la actividad actual. Los “teóricos de la memoria codificada” se ven obligados a concluir que el reconocimiento de cualquier estímulo implica el emparejamiento o la comparación sistemática del contenido del almacenamiento con los patrones sensoriales entrantes, con el fin de determinar si el estímulo actual se ha visto antes y determinar la respuesta adecuada del organismo. **Los teóricos de la tradición empirista, por otro lado, han combinado esencialmente la respuesta a la tercera pregunta con su respuesta a la segunda: dado que la información almacenada toma la forma de nuevas conexiones, o canales de transmisión en el sistema nervioso** [...] se deduce que los nuevos estímulos harán uso de estas nuevas vías que se han creado, activando automáticamente la respuesta adecuada sin requerir ningún proceso separado para su reconocimiento o identificación.

La teoría que se presenta aquí adopta la posición empirista o “conexionista” con respecto a estas cuestiones. La teoría ha sido desarrollada para un sistema nervioso hipotético, o máquina llamada perceptrón. El perceptrón está diseñado para ilustrar algunas de las propiedades fundamentales de los sistemas inteligentes en general, sin enredarse demasiado en las condiciones especiales, y con frecuencia desconocidas, que se dan en determinados organismos biológicos. La analogía entre el perceptrón y los sistemas biológicos debería ser evidente para el lector. Durante las últimas décadas, el desarrollo de la lógica simbólica, las computadoras digitales y la teoría de la conmutación ha impresionado a muchos teóricos con la similitud funcional entre una neurona y las unidades simples de encendido y apagado de las que se construyen las computadoras, y ha proporcionado los métodos analíticos necesarios para representar funciones lógicas muy complejas en términos de tales elementos. **El resultado ha sido una profusión de modelos cerebrales que equivalen simplemente a dispositivos lógicos para realizar algoritmos particulares** [...] Un número relativamente pequeño de teóricos, como Ashby (1) y von Neumann (17, 18), se han preocupado por los problemas de cómo una red neuronal imperfecta, que contiene muchas conexiones aleatorias, se puede hacer para realizar de manera confiable aquellas funciones que podrían ser representado por diagramas de cableado idealizados. **Desafortunadamente, el lenguaje de la lógica simbólica y el álgebra booleana es menos adecuado para tales investigaciones. La necesidad de un lenguaje adecuado para el análisis matemático de eventos en sistemas donde solo se puede caracterizar la organización general y se desconoce la estructura precisa, ha llevado al autor a formular el modelo actual en términos de teoría de la probabilidad más que de lógica simbólica.** [...]

La posición, elaborada por Hebb (7), Hayek (6), Uttley (16) y Ashby (1), en particular, sobre la que se basa la teoría del perceptrón, se puede resumir en los siguientes supuestos:

1. Las conexiones físicas del sistema nervioso que intervienen en el aprendizaje y el reconocimiento no son idénticas de un organismo a otro. Al nacer, la construcción de las redes más importantes es en gran parte aleatoria, sujeta a un número mínimo de limitaciones genéticas.
2. El sistema original de celdas conectadas es capaz de cierta plasticidad; después de un período de actividad neuronal, es probable que cambie la probabilidad de que un estímulo aplicado a un conjunto de células provoque una respuesta en otro conjunto, debido a algunos cambios relativamente duraderos en las propias neuronas.
3. A través de la exposición a una gran muestra de estímulos, aquellos que son más “similares” (en cierto sentido, que deben definirse en términos del sistema físico particular)

tenderán a formar vías hacia los mismos conjuntos de células que responden. Aquellos que son marcadamente “diferentes” tenderán a desarrollar conexiones con diferentes conjuntos de células que responden.

4. La aplicación de reforzamiento positivo y / o negativo (o estímulos que cumplen esta función) puede facilitar u obstaculizar cualquier formación de conexiones que esté en curso.

5. La similitud, en tal sistema, está representada en algún nivel del sistema nervioso por una tendencia de estímulos similares a activar los mismos conjuntos de células. La similitud no es un atributo necesario de clases particulares de estímulos formales o geométricos, sino que depende de la organización física del sistema de percepción, una organización que evoluciona a través de la interacción con un entorno dado. La estructura del sistema, así como la ecología del estímulo-ambiente, afectará y determinará en gran medida las clases de “cosas” en las que se divide el mundo perceptivo.

[...]”¹³⁵

El 13 de julio de 1958, el *New York Times* publicaría la historia que describe esta nueva máquina pensante llamada Perceptrón. El hombre que lo diseñó, Frank Rosenblatt, afirmó que eventualmente podría leer y escribir. Sería el “primer dispositivo en pensar como el cerebro humano ... Perceptrón cometerá errores al principio, pero se volverá más sabio a medida que gane experiencia”.¹³⁶ El Perceptrón fue una de las primeras redes neuronales utilizada en un equipo de cómputo (IBM 704). Su autor estaba convencido de que los algoritmos podían aprender como lo hacen los cerebros humanos, y su máquina hizo uso de esta arquitectura: como la red de neuronas del cerebro (la información viaja a través de capas de nodos interconectados).

En 1969, Marvin Minsky y Seymour Papert publicaron un libro llamado *Perceptrons* con el que pretendían cuestionar y desaparecer el uso de las redes neuronales del mapa de la IA. El libro influyó mucho (por el prestigio que tenían sus autores) y fue disminuyendo la investigación sobre redes neuronales en las universidades de todo el mundo desapareciendo casi por completo.¹³⁷

¹³⁵ F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain”, *Psychological review*, número 65, 6, 1958, pp. 386-408. https://pdfs.semanticscholar.org/5d11/aad09f65431b5d3cb1d85328743c9e53ba96.pdf?_ga=2.14359735.1221780561.1614451804-388228147.1609568917. [20 de julio de 2022]

¹³⁶ El texto original del artículo se puede consultar en el Anexo 4 de esta investigación.

¹³⁷ Fran, Ramírez, “Historia de la IA: Frank Rosenblatt y el Mark I Perceptrón, el primer ordenador fabricado específicamente para crear redes neuronales en 1957”, *LUCA*, 20 de julio de 2018. Consúltese en: <https://web.archive.org/web/20180722124753/https://data-speaks.luca-d3.com/2018/07/historia-de-la-ia-frank-roosenblatt-y-el.html#> [20 de julio de 2022]

En medio de este primer intento de sepultar las redes neuronales, aparece Geoffrey Hinton quien en 1978 obtiene su doctorado en Inteligencia Artificial en la Universidad de Edimburgo con el tema *Relaxation and its role in vision*.¹³⁸.

Hinton estudió a Kant, a Piaget y a Bartlett para explicar que la idea que entendemos el mundo asimilando a nuestros propios esquemas no es nueva. Utiliza a Minsky en sus disertaciones para contraargumentar sus propuestas e inventa el llamado “sistema SETTLE” para demostrar cómo la relajación puede usarse para controlar un sistema basado en datos que hace crecer una red relacional al notar cuándo las condiciones complejas se vuelven verdaderas y usar el encadenamiento hacia adelante. Una forma de organizar una búsqueda dentro de un sistema.

El capítulo sexto de su trabajo doctoral contiene la explicación de lo que él llama “la relajación”. Discute la relación entre el conocimiento almacenado y las representaciones creadas durante la percepción. Argumenta en contra de la idea de que la percepción es simplemente un proceso de hacer coincidir los datos con los modelos almacenados, y así preparar el terreno para el sistema SETTLE. Aquí cambia el enfoque tradicional de la IA basado en el razonamiento lógico y la correspondiente introducción de un enorme número de datos, reglas de inferencia y heurísticas para determinar cuándo aplicarlos hacia una visión contrapuesta basada en la biología y el intentar imitar las redes neuronales del cerebro humano, en el que la percepción a través de los sentidos, van aprendiendo y generando conocimiento a partir de la experiencia -ensayo/error- (así no hay que especificar todo el conocimiento).

Geoffrey Hinton¹³⁹, es el creador del algoritmo *de Backpropagation* que permite detectar aciertos y errores partiendo de un resultado final y volviendo sobre los pasos que se siguieron previamente. También desarrolló el algoritmo que creó

¹³⁸ G. E. Hinton, *Relaxation and its role in vision*. Tesis de doctorado, Universidad de Edimburgo, Edimburgo, 1978. Consúltese el texto original en: <https://era.ed.ac.uk/handle/1842/8121> [20 de julio de 2022]

¹³⁹ Nació en el Reino Unido. Su tatarabuelo paterno fue el matemático George Boole. Hinton estaba obsesionado con comprender la mente y sintió que, para hacerlo, uno debe comprender el cerebro. Estudió Psicología Experimental en Cambridge (1970). Obtuvo un doctorado en Inteligencia Artificial en la Universidad de Edimburgo (1978). Actualmente es profesor emérito en el Departamento de Ciencias de la Computación de la Universidad de Toronto (Canadá). Desde 2013 es vicepresidente en el área de Ingeniería de Google. Para más datos consúltese: <https://www.cs.toronto.edu/~hinton/fullcv.pdf> [20 de julio de 2022]

las redes neuronales profundas (RNP), el cual ha permitido una revolución total de la IA en sectores relacionados con el reconocimiento de imágenes y el procesamiento del lenguaje natural, vehículos sin conductor e interpretación de imágenes médicas para predecir el cáncer.

Después de los éxitos de la década de 1980 que tuvo Hinton, hubo un período difícil en el que muchos científicos sugirieron dejar el uso de las redes neuronales, Hinton fue uno de los pocos que siguió adelante, a pesar de la escasez de recursos (deja Reino Unido y se mueve a Toronto para recibir una modesta subvención).

En 2006, Hinton hizo un gran avance rebautizándolas como redes neuronales de aprendizaje profundo, cambiando esto todo el modelo tradicional de la IA en todas las tareas relacionadas con el reconocimiento de imágenes, voz, texto. En todas estas actividades IBM, Amazon, Google, Facebook, Microsoft están compitiendo por mejorar cotidianamente la tecnología y explotarla al máximo.

Vale la pena aclarar algunos conceptos para los fines de nuestra investigación. El actual DL describe a “una familia de modelos de Redes Neuronales Profundas que tienen múltiples capas de procesamiento de información organizadas en red... Una red neuronal de este tipo está inspirada en la estructura del cerebro humano.”¹⁴⁰ Y una Red Neuronal Artificial consiste en una red de

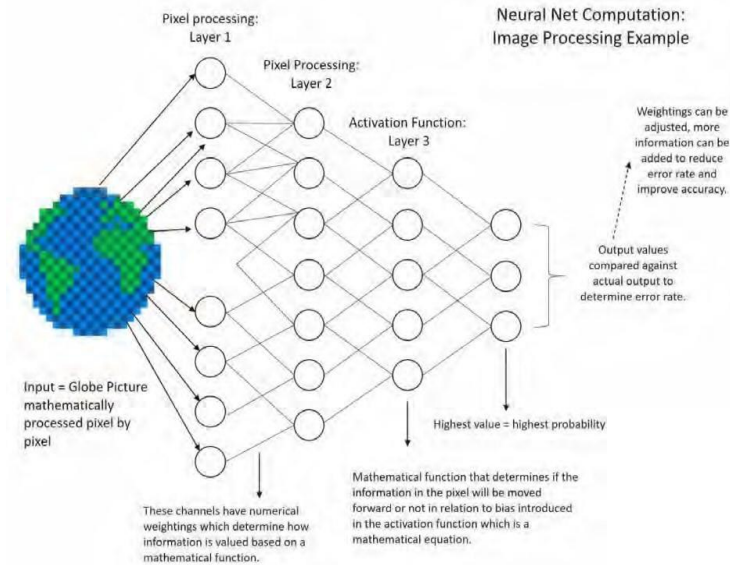
¹⁴⁰ Cfr. John D., Kelleher, *Deep Learning, The MIT Press Essential Knowledge series*, Cambridge, The MIT Press, Massachusetts, 2019, pp. 65-67. Kelleher comenta: “El cerebro humano está compuesto por una gran cantidad de células nerviosas, llamadas neuronas. De hecho, algunas estimaciones sitúan el número de neuronas en el cerebro humano en cien mil millones... Las neuronas tienen una estructura simple de tres partes que consta de: un cuerpo celular, un conjunto de fibras llamadas dendritas y una sola fibra larga llamada axón.. Las dendritas y el axón provienen del cuerpo celular, y las dendritas de una neurona están conectados a los axones de otras neuronas. Las dendritas actúan como canales de entrada a la neurona y reciben señales enviadas desde otras neuronas a lo largo de sus axones. El axón actúa como el canal de salida de una neurona, por lo que otras neuronas, cuyas dendritas están conectadas al axón, reciben las señales enviadas a lo largo del axón como entradas. Las neuronas funcionan de una manera muy simple. Si los estímulos entrantes son lo suficientemente fuertes, la neurona transmite un pulso eléctrico, llamado potencial de acción, a lo largo de su axón a las otras neuronas que están conectadas a ella. Entonces, una neurona actúa como un interruptor de todo o nada, que recibe un conjunto de entradas y emite un potencial de acción o no produce ninguna salida. Esta explicación del cerebro humano es una simplificación significativa de la realidad biológica, pero captura los puntos principales necesarios para comprender la analogía entre la estructura del cerebro humano y los modelos computacionales llamados redes neuronales. Estos puntos de analogía son: (1) el cerebro está compuesto por un gran número de unidades interconectadas y simples llamadas neuronas; (2) el funcionamiento del cerebro puede entenderse como el procesamiento de información, codificada como señales eléctricas altas o bajas, o potenciales de activación, que se propagan a través de la red de neuronas; y (3) cada

unidades simples de procesamiento de información... El poder de las redes neuronales para modelar relaciones complejas no es el resultado de modelos matemáticos complejos, sino que surge de las interacciones entre un gran conjunto de neuronas simples. Las redes de aprendizaje profundo son redes neuronales que tienen muchas capas ocultas de neuronas.¹⁴¹ La siguiente imagen nos muestra una descripción de un modelo de redes neuronales profundas modernas.

neurona recibe un conjunto de estímulos de sus vecinas y asigna estas entradas a una salida de valor alto o bajo. Todos los modelos computacionales de redes neuronales tienen estas características." [La traducción es propia]

¹⁴¹ Ibidem., p. 68-69.

Descripción del procesamiento de datos en redes neuronales



Fuente: Lydia Kostopoulos, Decoupling Human Characteristics from algorithmic capabilities, The IEEE Standards Association, January 2021.

En 2011, Jeff Dean, ingeniero de Google se encontró con Andrew Ng, profesor de informática e investigador visitante de la Universidad de Stanford. Dean se sorprendió al saber que Ng estaba usando redes neuronales. Formaron un equipo que terminó en el proyecto "Google X", un laboratorio semisecreto para proyectos ambiciosos y de largo alcance. Diseñaron una red neuronal masiva (mil millones de conexiones distribuidas en 16,000 procesadores de computadora) y la alimentaron con una base de datos de 10 millones de capturas de fotogramas aleatorios de YouTube. Las imágenes no estaban etiquetadas, es decir el modelo de computadora recibió un mar de píxeles sin ninguna otra información.

¿Podría aprender a detectar objetos sin ayuda humana? Los investigadores encontraron como resultado de la revisión tres imágenes borrosas: un rostro humano, un cuerpo humano y un gato. "El modelo en realidad estaba recogiendo patrones que tú, como ser humano, también dirías que son importantes, sin que nunca te lo digan", afirma Jeff Dean. El equipo salió de los laboratorios de Google X y silenciosamente comenzó a reunir un pequeño grupo de especialistas en redes neuronales de todas partes del mundo:

- a) En 2011, Navdeep Jaitly, un estudiante de doctorado con Hinton se convirtió en pasante de Google. Se le pidió que modificara los algoritmos de reconocimiento de voz de Google, y respondió sugiriendo que destruyeran la mitad de su sistema y lo reemplazaran con una red neuronal. El programa de Jaitly superó a los sistemas que se habían perfeccionado durante años, y los resultados de su trabajo se abrieron camino en Android, el sistema operativo móvil de Google, un golpe poco común para un pasante. Jaitly ahora trabaja en Google Brain como científico investigador a tiempo completo.
- b) En 2012, Hinton y dos de sus otros estudiantes de la U of T, Alex Krizhevsky e Ilya Sutskever, participaron en un concurso de reconocimiento de imágenes. Los equipos que compiten construyen algoritmos de visión por computadora que aprenden a identificar objetos en millones de imágenes; las victorias más precisas. El modelo U de T tomó la tasa de error de los algoritmos de mejor desempeño hasta la fecha, y la tasa de error del humano promedio, y partió la diferencia a la mitad como una ramita seca. El trío creó una empresa y Google la adquirió.
- c) En 2013, investigadores de una pequeña empresa británica llamada *DeepMind Technologies* publicaron una preimpresión de un artículo de investigación que mostraba cómo había enseñado a una red neuronal a jugar y vencer a los juegos de Atari¹⁴².
- d) En 2014 Google compra *DeepMind*. Facebook contrató al pionero de la red neuronal Yann LeCun para dirigir un laboratorio de investigación de inteligencia artificial a principios de ese invierno. Baidu, el gigante de las búsquedas chino, incorporó a Andrew Ng después de que dejó Google.¹⁴³

¹⁴² Kate, Allen, "Computer beats Atari in artificial intelligence breakthrough", *Toronto Star*, febrero 25, 2015. Consúltase en: <https://www.thestar.com/news/world/2015/02/25/computer-beats-atari-in-artificial-intelligence-breakthrough.html> [20 de julio de 2022]

¹⁴³ Kate, Allen, "How a Toronto professor's research revolutionized artificial intelligence", *Toronto Star*, abril 17, 2015. Consúltase en <https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html> [20 de julio de 2022]

En palabras de Hinton:

“La máquina que mejor aprende de todas las que conocemos es el cerebro humano. Y la forma en que funciona es que aprende cambiando la fuerza de las conexiones entre los miles de millones de neuronas que tiene. Así que una manera de hacer que un ordenador aprenda es hacer que el ordenador simule ser un montón de neuronas e intente encontrar una regla para cambiar la fuerza de la conexión entre las neuronas para que aprenda igual que lo hace el cerebro. La idea es darle muchos ejemplos de las entradas y salidas que te gustaría que produjera y luego se le cambia la fuerza de conexión para que al mostrarle una entrada obtenga la respuesta correcta. ***Mi investigación se ha centrado, principalmente, en intentar averiguar cuáles son las reglas que nos permiten reforzar las conexiones de la red neuronal, para que la máquina sea capaz de aprender. Y eso nos conduce a un nuevo tipo de inteligencia artificial donde ya no se programa el conocimiento; el ordenador aprende por sí mismo, por su propia experiencia.***”¹⁴⁴

Yann LeCun, Yoshua Bengio y Geoffrey Hinton¹⁴⁵, el 27 de mayo de 2015, dan a conocer parte de sus avances científicos con el artículo “*Deep Learning*”. El aprendizaje profundo -comentan los autores- permite que los modelos computacionales que están compuestos por múltiples capas de procesamiento aprendan representaciones de datos con múltiples niveles de abstracción. Estos métodos han mejorado drásticamente el estado de la técnica en reconocimiento de voz, reconocimiento de objetos visuales, detección de objetos y muchos otros dominios, como el descubrimiento de fármacos y la genómica.

El aprendizaje profundo descubre una estructura intrincada en grandes conjuntos de datos mediante el uso del algoritmo de retropropagación para indicar cómo una máquina debe cambiar sus parámetros internos que se utilizan para calcular la representación en cada capa a partir de la representación en la capa anterior. Las redes convolucionales profundas han producido avances en el procesamiento de imágenes, video, voz y audio, mientras que las redes recurrentes han arrojado luz sobre datos secuenciales como texto y voz.”¹⁴⁶

¹⁴⁴ Premios frontera del conocimiento 2016, “La Fundación BBVA premia a Geoffrey Hinton artífice de las primeras máquinas capaces de aprender como el cerebro humano”, Fundación BBVA, Salamanca, 17 de enero 2017. Consúltase en <https://www.premiosfronterasdelconocimiento.es/noticias/la-fundacion-bbva-premia-al-artifice-las-primeras-maquinas-capaces-aprender-cerebro-humano/> [20 de julio de 2022]

¹⁴⁵ Y., LeCun, Y., Bengio, y G., Hinton, “*Deep Learning*”, *Nature*, número 521, 2015, pp. 436-444. Consúltase en: <https://doi.org/10.1038/nature14539> [20 de julio de 2022]

¹⁴⁶ Ibidem., p. 1. [La traducción es propia]

El futuro de la IA está inclinándose hacia el uso del aprendizaje no supervisado al tener éste un efecto catalizador reviviendo el aprendizaje profundo. Se pronostica que el aprendizaje no supervisado será mucho más importante a largo plazo. En el artículo los autores concluyen lo siguiente:

“El aprendizaje humano y animal en gran parte no está supervisado: descubrimos la estructura del mundo observándolo, no diciéndonos el nombre de cada objeto. La visión humana es un proceso activo que muestrea secuencialmente la matriz óptica de una manera inteligente y específica de la tarea utilizando una fovea pequeña de alta resolución con un marco grande de baja resolución. **Esperamos que gran parte del progreso futuro en la visión provenga de sistemas entrenados de principio a fin y que combinen ConvNets con RNN que utilizan el aprendizaje por refuerzo para decidir dónde buscar...** La comprensión del lenguaje natural es otra área en la que el **aprendizaje profundo** está preparado para tener un gran impacto en los próximos años. Esperamos que los sistemas que usan RNN para comprender oraciones o documentos completos se vuelvan mucho mejores cuando aprendan estrategias para atender selectivamente una parte a la vez. En última instancia, los principales avances en inteligencia artificial se producirán a través de sistemas que **combinan el aprendizaje de la representación con el razonamiento complejo**. Aunque el aprendizaje profundo y el razonamiento simple se han utilizado para el reconocimiento de voz y escritura a mano durante mucho tiempo, se necesitan nuevos paradigmas para reemplazar la manipulación basada **en reglas de expresiones simbólicas por operaciones en vectores grandes**.¹⁴⁷

En 2019, Yoshua Bengio, Geoffrey Hinton y Yann LeCun, obtuvieron el Premio Turing por sus aportaciones al campo del *Deep Learning*. LeCun comenta que desde mediados de la década de los años 1990 hasta mediados de la década del

¹⁴⁷ El texto original de sus conclusions es el siguiente: “Unsupervised learning had a catalytic effect in reviving interest in *Deep Learning* but has since been overshadowed by the successes of purely supervised learning. Although we have not focused on it in this Review, we expect unsupervised learning to become far more important in the longer term. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object. Human vision is an active process that sequentially samples the optic array in an intelligent, task-specific way using a small, high-resolution fovea with a large, low-resolution surround. We expect much of the future progress in vision to come from systems that are trained end-to-end and combine ConvNets with RNNs that use reinforcement learning to decide where to look. Systems combining *Deep Learning* and reinforcement learning are in their infancy, but they already outperform passive vision systems at classification tasks and produce impressive results in learning to play many different video games. Natural language understanding is another area in which *Deep Learning* is poised to make a large impact over the next few years. We expect systems that use RNNs to understand sentences or whole documents will become much better when they learn strategies for selectively attending to one part at a time. Ultimately, major progress in artificial intelligence will come about through systems that combine representation learning with complex reasoning. Although *Deep Learning* and simple reasoning have been used for speech and handwriting recognition for a long time, new paradigms are needed to replace rule-based manipulation of symbolic expressions by operations on large vectors.”. Ibidem., p. 442.

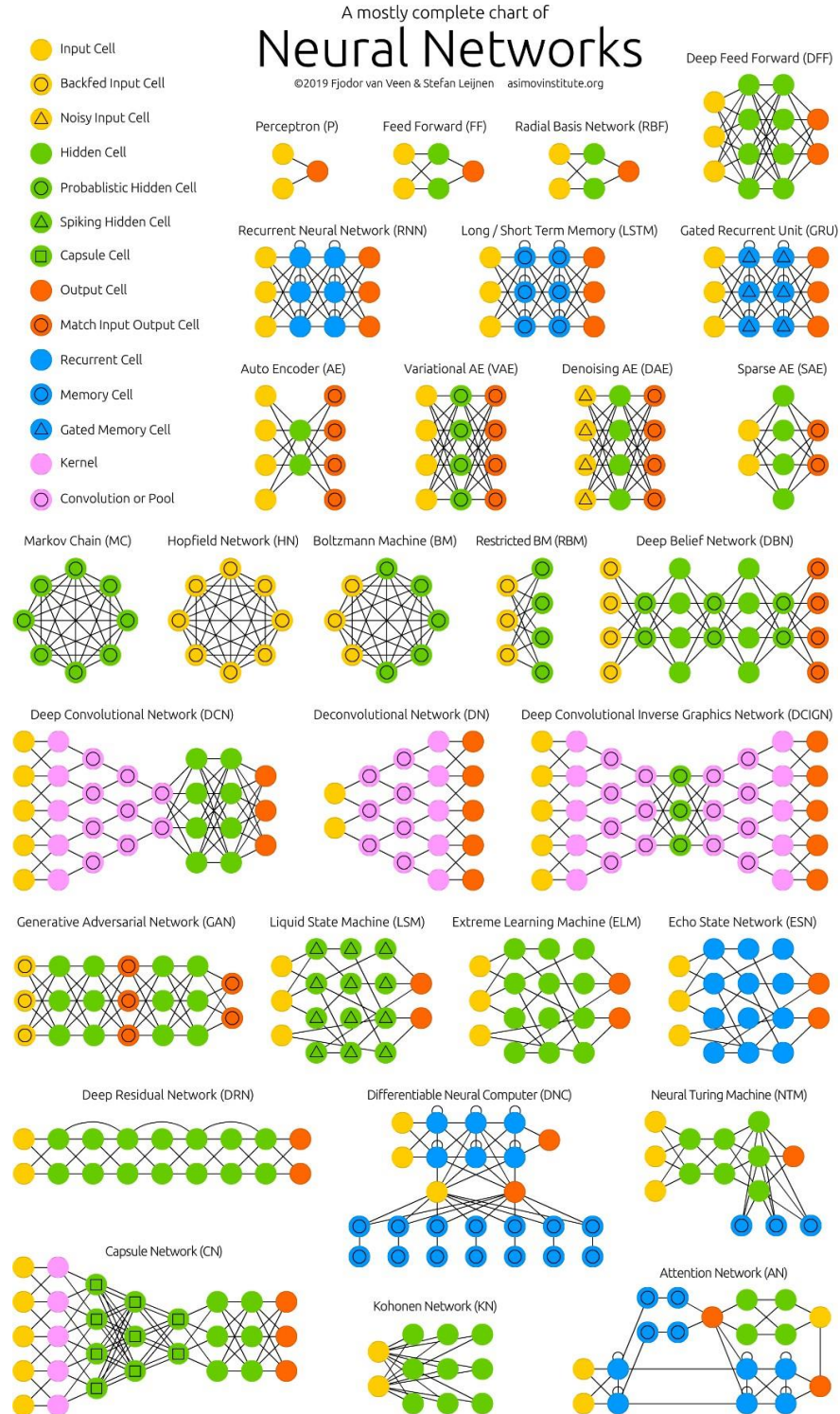
2000 la comunidad científica había perdido interés en este tema, pero lograron organizar reuniones y talleres regulares de verano, creándose una pequeña comunidad que protagonizó un cambio fundamental a partir de 2012-2013.

El punto de quiebre de esa mala imagen que se tenía sobre el uso del *Deep Learning* llega cuando Hinton ganó el concurso *ImageNet* (2012) gracias a una red neuronal convolucional llamada AlexNet, que logró superar a todos sus competidores por una amplia diferencia.¹⁴⁸ A partir de ahí, el mundo de la IA está dejando la precisión lógica por la percepción biológica.

La siguiente imagen se denomina *The neural network zoo* y es tomada de *The Asimov Institute*, organización independiente que se dio a la tarea de compilar los principales tipos de algoritmos basados en redes neuronales profundas y, también, hace una breve descripción de cada una que se puede consultar en el Anexo 6 de este apartado:

¹⁴⁸ Marcos Merino, “El Premio Turing, para tres de los principales responsables del actual auge de la inteligencia artificial”, Xataka, 29 de marzo de 2019. Consúltase en: <https://www.xataka.com/inteligencia-artificial/premio-turing-para-tres-principales-responsables-actual-auge-inteligencia-artificial#:~:text=Geoffrey%20Hinton,-Geoffrey%20Hinton%2C%20brit%C3%A1nico&text=En%201986%20invent%C3%B3%20junto%20a,supervisado%20de%20transformaciones%20de%20imagen%22>. [20 de julio de 2022]

Tipos de algoritmos basados en redes neuronales profundas



Fuente: The Asimov Institute <https://www.asimovinstitute.org/neural-network-zoo/>
[15 de julio de 2022]

Para finalizar, esta sección muestra algunos usos que se le dan a los modelos de ML – DL tanto en la vida cotidiana como en las actividades científicas. Las funciones del ML – DL son de uso común o cotidiano y que las empresas, gobiernos y sociedad están siendo beneficiados. Hemos dividido en dos estos beneficios: ML – DL en el quehacer cotidiano (1) y ML – DL en las actividades científicas:

1. ML - DL en el quehacer cotidiano

Sistemas de recomendación de productos o servicios: Se encuentran entre las aplicaciones de aprendizaje automático más reconocidas. No hay necesidad de saber utilizar o conocer la tecnología para que el usuario utilice ML. Utilizan patrones de consumo y preferencias que el usuario va ingresando para predecir qué productos o servicios pueden ser deseables para próximas consultas. Se utiliza en una variedad de empresas minoristas. También se usa para promover determinados contenidos a los usuarios de las redes sociales, noticias o uso de determinados sitios de internet.

Mapas: Las aplicaciones de mapeo utilizan datos históricos e información de tráfico en tiempo real para planificar la mejor ruta que deben tomar los usuarios de la carretera. Si una calle está muy transitada, estas aplicaciones pueden redireccionar a las personas a una ruta con menor tránsito. También utilizan datos GPS de los teléfonos móviles de los conductores para proporcionar información en tiempo real sobre la velocidad y el tráfico.

Organización de la información a través de motores de búsqueda y filtrado de spam: Ayuda a proporcionar los resultados de las consultas ingresadas en los motores de búsqueda de Internet. Se ingresa la palabra y el sistema hace la búsqueda dando como resultado palabras y frases que tienen significados iguales o muy similares y usan esta información para predecir las páginas web adecuadas para responder a la consulta. También existen los sistemas de detección de spam para el filtrado de correos electrónicos. En esta aplicación, el sistema se capacita utilizando una muestra de documentos, que se clasifican como spam y no spam, para distinguir entre correos electrónicos y dirigirlos a las carpetas correctas. El sistema va aprendiendo cómo la presencia de palabras o los nombres de

remitentes, y otras características, se relacionan con si el correo electrónico es o no un spam.

Reconocimiento de voz y respuesta por asistentes personales virtuales: Los sistemas de reconocimiento de voz y procesamiento del lenguaje natural pueden coincidir con los patrones de sonidos producidos en el habla humana a palabras o frases que ya han encontrado, distinguiendo entre las diferentes huellas de audio de estos sonidos. Una vez identificadas las palabras utilizadas, pueden traducirlas a texto o ejecutar comandos. Los encontramos en muchos teléfonos inteligentes y otros dispositivos como bocinas, computadoras, tabletas.

Visión por computadora para etiquetado de fotos y reconocimiento de escritura a mano: Las computadoras puedan detectar y analizar imágenes y asociar información numérica o simbólica a ellas. En las redes sociales, el reconocimiento de imágenes se puede utilizar para etiquetar objetos o personas en fotos que se han subido a sitios web. También se pueden utilizar sistemas de reconocimiento de imágenes para material escrito a mano que es escaneado. Ayuda a los usuarios a etiquetar las fotos de amigos y familiares en las redes sociales. Le da al usuario diferentes formas de clasificación.

Traducción automática de texto a diferentes idiomas: Se puede convertir automáticamente el texto o el habla de un idioma a otro. La traducción automática se utiliza en aplicaciones de traducción específicas para teléfonos móviles, medios sociales y tradicionales, y en organismos internacionales como la ONU, OEA, UE, que necesitan reproducir documentos en una gran cantidad de idiomas.

Detección de patrones en el sector financiero: Se utiliza para identificar patrones en movimientos financieros cuyos montos, número de operaciones, lugar de origen y destino de operaciones generen alertas para prevención de lavado de dinero que para los seres humanos implican una labor de trabajo enorme. También existen sistemas para la detección de fraude asociados con el uso de tarjetas de crédito u otros sistemas de pago. Puede detectar operaciones inusuales en determinados usuarios para que el oficial de cumplimiento o el área de prevención de fraudes levante una alerta y se verifique al usuario.

Uso de *bots*: El término *bot* se utiliza para referirse a un agente implementado en software. Estos agentes pueden no tener una manifestación física, pero pueden operar de forma autónoma en el mundo virtual y guiar a las personas en la elección de determinados productos o servicios y en actividades de atención a clientes vía telefónica o por internet.

Políticas públicas y política criminal: En algunas ciudades, como Nueva York, se han utilizado las aplicaciones de ML para definir políticas públicas o la política criminal basándose en las capacidades creadas con la visión por computadora y el procesamiento del lenguaje.

2. ML - DL en las labores científicas

Estadísticas, ciencia de datos, robótica e inteligencia artificial: El aprendizaje automático está estrechamente relacionado con los campos de la estadística y la ciencia de datos, que proporcionan una variedad de herramientas y métodos para el análisis de la información que se utiliza para nutrir a los sistemas y la inferencia de esa información para sacar conclusiones que ayuden a interpretar el contenido.

También está relacionado con la robótica y la automatización inteligente. Los sistemas robóticos contribuyen a aplicaciones como vehículos autónomos y drones.

Ciencias de la salud: El uso de ML puede ayudar a diagnosticar problemas de salud mental o Parkinson al analizar los patrones del habla e identificar las características de estas enfermedades. En el tema del cáncer, ha identificado y medido 6,642 características en el cáncer (en el pasado, un ser humano evaluaba tres características específicas del cáncer de mama mirando imágenes a través de un microscopio), con un mejor desempeño que los analistas humanos, pero también descubriendo nuevas características. Las investigaciones sobre el genoma humano son hechas con ML.

El ML ha sido utilizado para la solución de varios tipos de problemas. ¿Cómo se decide qué algoritmo utilizar? Saber identificar el tipo de problema y el tipo de algoritmo que se puede aplicar es muy importante para lograr los resultados esperados por los usuarios, la empresa o el gobierno. Algunos ejemplos más comunes son los siguientes:

Selección de algoritmos según el problema a resolver (reglas básicas)

Preguntarse	Ejemplos de problemas	Tipo de algoritmo
¿A qué categoría pertenece este tipo de datos?	Diagnóstico médico: ¿este tejido muestra signos de enfermedad? Banca: ¿esta transacción es fraudulenta? Visión por computadora: ¿qué tipo de objeto hay en esta imagen? ¿Es una persona? ¿Es un edificio?	Clasificación
Dado conjunto de datos, ¿cuál es el valor probable de una cantidad en particular?	Finanzas: ¿cuál será el valor de esta acción mañana? Vivienda: ¿cuál sería el precio de esta casa si se vendiera hoy? Calidad alimentaria: ¿cuántos días antes de que madure esta fresa? Procesamiento de imágenes: ¿cuántos años tiene la persona de esta foto?	Regresión
¿Qué tipo de datos son similares entre sí?	Comercio electrónico: ¿qué clientes muestran un comportamiento similar entre sí, ¿cómo se agrupan? Transmisión de video: ¿Cuáles son los diferentes tipos de géneros de video en nuestro catálogo y qué videos pertenecen al mismo género?	Agrupación (Clustering)
¿Cuáles son las características más importantes de estos datos y cómo se pueden resumir?	Comercio electrónico: ¿qué combinaciones de funciones nos permiten resumir el comportamiento de nuestros clientes? Biología molecular: ¿cómo pueden los científicos resumir el comportamiento de los 20.000 genes humanos en un tejido enfermo en particular?	Reducción de dimensionalidad
¿Cómo se pueden combinar los datos etiquetados y no etiquetados?	Visión por computadora: ¿cómo se puede desarrollar la detección de objetos con solo un pequeño conjunto de datos de entrenamiento? Descubrimiento de fármacos: ¿cuál de los millones de posibles fármacos podría ser eficaz contra enfermedades, dado que hasta ahora solo hemos probado unos pocos?	Aprendizaje semi-supervisado
¿Qué acciones lograrán con mayor eficacia el punto final deseado?	Robots: ¿cómo puede un robot moverse por su entorno? Juegos: ¿Qué movimientos fueron importantes para ayudar a la computadora a ganar un juego en particular?	Aprendizaje por refuerzo

Fuente: El autor con base en: *The Royal Society, op. cit., p. 29.*

Elegir el algoritmo correcto puede ser bastante abrumador: hay una variedad amplia de algoritmos de aprendizaje automático supervisados, no supervisados y por refuerzo; cada uno tiene un enfoque diferente para el aprendizaje. No existe el mejor método o una única ruta de elección. Encontrar el algoritmo correcto es, en parte, un ejercicio de “aproximaciones sucesivas” (ensayo y error); incluso los científicos de datos con mucha experiencia no pueden saber si un algoritmo funcionará correctamente hasta que empieza las pruebas en la etapa de producción. Pero la selección del algoritmo también depende del tamaño y el tipo de datos con los que se está trabajando, la información que desea obtener de los datos y para qué se utilizará esa información¹⁴⁹.

En 2006, la Conferencia Internacional de Minería de Datos (*International Conference on Data Mining, ICDM*) del IEEE (*Institute of Electrical and Electronics Engineers*) ubicó 10 algoritmos relevantes para la comunidad científica:

1. C4.5,
2. k-Means,
3. SVM,
4. Apriori,
5. EM,
6. PageRank,
7. AdaBoost,
8. KNN,
9. Naive Bayes;
10. CARRO.

¹⁴⁹ Respecto la evolución de este tema consúltese: Cfr. Kubat, Miroslav, *An Introduction to Machine Learning*, Springer, Florida, 2017. Kubat comenta que se pueden encontrar interesantes datos sobre la cuestión de la representación de conceptos o clases en algunos filósofos como John Stuart Mill (exploró conceptos relacionados con la representación probabilística) y William Whewell (defendían representaciones prototípicas). Existe también el libro *Introduction to Evolutionary Computing*, de A.E. Eiben y J.E. Smith (Springer, 2015) en el que sus autores dedican la segunda parte en los problemas de cómo representar, elegir y diseñar algoritmos “evolutivos” para diferentes actividades y la recopilación de soluciones al control de parámetros. El libro ofrece una introducción a la computación evolutiva (CE), descripciones de variantes populares de algoritmos evolutivos (EA), discusiones sobre cuestiones metodológicas y técnicas de CE particulares.

Como primer paso en el proceso de identificación de estos algoritmos, en 2006, se invitó a los ganadores del premio ACM KDD *Innovation Award* y del premio IEEE ICDM *Research Contributions Award* para que cada uno nominara hasta 10 algoritmos más conocidos en minería de datos. Se les pidió que proporcionara la siguiente información: (a) el nombre del algoritmo, (b) una breve justificación y (c) una referencia de publicación representativa. Cada algoritmo debía haber sido ampliamente citado y utilizado por otros investigadores en el campo, y las nominaciones deberían tener una representación razonable de las diferentes áreas de la minería de datos.¹⁵⁰ Se verificaron las nominaciones a través de Google Scholar, eliminándose aquellas que no tenían al menos 50 citas. Las nominaciones restantes se organizaron en 10 temas:

1. análisis de asociación,
2. clasificación,
3. agrupamiento,
4. aprendizaje estadístico,
5. ensacado y refuerzo,
6. patrones secuenciales,
7. minería integrada,
8. conjuntos aproximados,
9. minería de enlaces
10. minería de gráficos.

En el tercer paso del proceso de identificación se invitó a los miembros del Comité de Programa de KDD-06 (la Conferencia Internacional ACM SIG-KDD sobre Descubrimiento de Conocimiento y Minería de Datos), ICDM '06 (la Conferencia Internacional IEEE de 2006 sobre Minería de Datos) y SDM '06 (la Conferencia Internacional SIAM 2006 - Conferencia nacional sobre minería de datos), así como los ganadores del Premio a la innovación ACM KDD y del Premio a las contribuciones a la investigación IEEE ICDM por cada voto de hasta 10 algoritmos conocidos de la lista de candidatos de los algoritmos. Los resultados de la votación

¹⁵⁰ Wu, X., Kumar, V., Ross Quinlan, J. et al. (2008). "Top 10 algorithms in data mining", *Know Inf Syst* 14, 1–37. <https://doi.org/10.1007/s10115-007-0114-2> [20 julio de 2022]

de fueron concentrados en el texto *Top 10 algorithms in data mining*. (Wu, X., Kumar, V., Ross Quinlan, J. *et al.*, 2008)¹⁵¹

Haciendo una revisión a MathWorks¹⁵², se encontró que el método de aprendizaje supervisado¹⁵³ procesa los datos mediante los siguientes tipos de algoritmos:

- Para valores de respuesta categóricos, en los que los datos se pueden separar en clases específicas siendo los algoritmos más comunes (clasificación):
 - ✓ Máquinas de vector soporte (Support Vector Machine, SVM)
 - ✓ Redes neuronales
 - ✓ Clasificador *Naïve Bayes* (NBC)
 - ✓ Árbol de decisiones
 - ✓ Análisis discriminante
 - ✓ Vecinos más cercanos (*K Nearest Neighbours K-NN*)
- Para valores de respuesta continua siendo los algoritmos más comunes (regresión):
 - ✓ Regresión lineal
 - ✓ Regresión no lineal
 - ✓ Modelos lineales generalizados
 - ✓ Árboles de decisión
 - ✓ Redes neuronales

A la vez, se encontró que el método de aprendizaje no supervisado¹⁵⁴ utiliza el análisis de clústeres para hacer análisis de datos exploratorios y buscar patrones

¹⁵¹ Este es un ejemplo de algoritmos utilizados para ciencia de datos, pero dependiendo el acto, el sector de bienes o servicios que se desee automatizar será el tipo de algoritmo que se deba utilizar.

¹⁵² MathWorks fue fundada en 1984 por Jack Little y Cleve Moler, que identificaron la necesidad de los ingenieros y los científicos de disponer de entornos de cálculo más potentes y productivos que los proporcionados por lenguajes de programación como Fortran y C. Combinaron su experiencia en matemáticas, Ingeniería e informática para desarrollar MATLAB. Consúltase: https://la.mathworks.com/company.html?s_tid=hp_ff_a_company [20 julio de 2022]

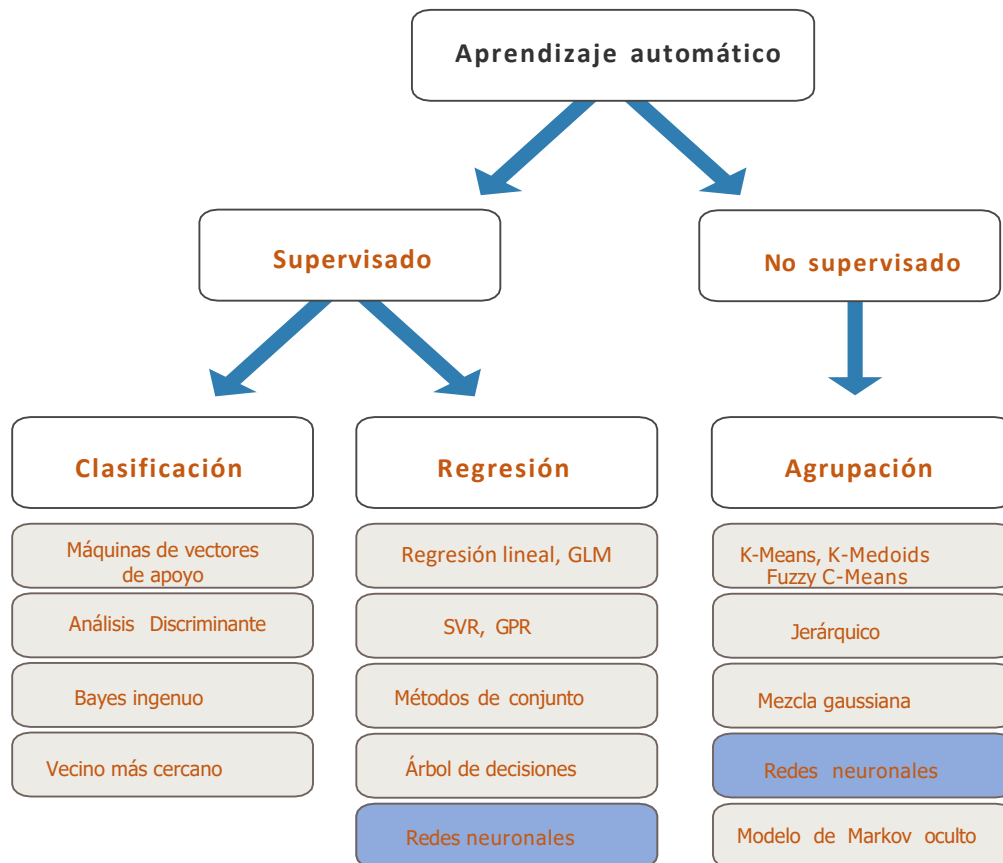
¹⁵³ El aprendizaje supervisado se emplea en aplicaciones financieras para la calificación crediticia, el *trading* algorítmico y la calificación de bonos; en aplicaciones médicas para la detección de tumores y el descubrimiento de fármacos; en aplicaciones energéticas para la predicción de carga y precio; y en aplicaciones de reconocimiento de patrones para el habla y las imágenes. Consúltase: <https://la.mathworks.com/discovery/supervised-learning.html> [20 julio de 2022]

¹⁵⁴ Los métodos de aprendizaje no supervisados se utilizan en la bioinformática para el análisis de secuencias y la agrupación genética; en minería de datos para minería de secuencias y patrones;

ocultos o agrupación en datos. Los algoritmos de agrupación en clústeres comunes incluyen:

- Clústeres jerárquicos: crea una jerarquía multinivel de clústeres mediante la creación de un árbol de clústeres.
- *k-Means clustering*: divide los datos en k clústeres distintos en función de la distancia al centroide de un clúster.
- Modelos de mezcla gaussiana: los modelos se agrupan como una mezcla de componentes de densidad normal multivariados.
- Mapas auto organizadores: utiliza redes neuronales que aprenden la topología y distribución de los datos.
- Modelos ocultos de Markov: utiliza datos observados para recuperar la secuencia de estados.

Ejemplos de algoritmos y su relación con el ML



en imágenes médicas para la segmentación de imágenes; y en visión computarizada para el reconocimiento de objetos. Consúltese: <https://la.mathworks.com/discovery/unsupervised-learning.html> [20 julio de 2022]

Consideramos relevante reiterar que, a pesar del uso generalizado del ML, existen mitos sobre lo que puede hacer y lo que no puede hacer el ML - DL. Cynthia Rudin y David Carlson ubican “10 secretos” relevantes que se deben tener en cuenta para evitar caer en esos mitos del ML - DL¹⁵⁵

- 1) Hay al menos cuatro familias principales para el aprendizaje supervisado: métodos de modelado lógico, métodos de combinación lineal, métodos de razonamiento basado en casos y métodos de resumen iterativo.
- 2) Para muchos dominios de aplicaciones, casi todos los métodos de aprendizaje automático funcionan de manera similar (con algunas salvedades). Los métodos de aprendizaje profundo, que son la técnica líder para los problemas de visión de las computadoras, no mantienen una ventaja sobre otros métodos para la mayoría de los problemas (y hay razones para ello).
- 3) Las redes neuronales son difíciles de entrenar y, a menudo, suceden cosas raras cuando intentas entrenarlas.
- 4) Si no utiliza un modelo interpretable, puede cometer graves errores.
- 5) Las explicaciones pueden ser engañosas y no puedes confiar en ellas.
- 6) Casi siempre puede encontrar un modelo preciso pero interpretable, incluso para redes neuronales profundas.
- 7) Se deben incorporar propiedades especiales como la toma de decisiones o la solidez, no ocurren por sí solas.
- 8) La inferencia causal es diferente a la predicción (la correlación no es causalidad).
- 9) Siempre existe un método para controlar las redes neuronales profundas.
- 10) Es un mito que la inteligencia artificial puede hacer cualquier cosa.

Estas recomendaciones si se usan de manera correcta y con prudencia, pueden hacer eficientes un sinnúmero de actividades (sector público, sector privado, vida de las personas) pero si se usan de manera incorrecta, pueden causar daños a Derechos Humanos como los ocurridos en cuestiones relacionadas con

¹⁵⁵ C., Rudin, y D.E., Carlson, “The Secrets of *Machine Learning*: Ten Things You Wish You Had Known Earlier to be More Effective at Data Analysis”, *ArXiv*, *abs/1906.01998*, 2019. [20 de julio de 2022]

justicia penal.¹⁵⁶ Dada la gran cantidad de publicidad que hay actualmente en torno al ML - DL, junto con el asombroso progreso, pero también los diversos fracasos, se ha vuelto difícil separar toda esta publicidad de la realidad y los abogados nos dejamos llevar por lo que sueña en el foro jurídico más que “por lo que estudia el foro jurídico”.

Se ha visto también que los modelos de ML tienen sesgos al momento de relacionar grupos de personas y la evaluación de sus currículos cuando sus diseñadores no tienen en cuenta la forma en que el algoritmo se utilizará en la práctica.¹⁵⁷ Y los algoritmos de caja negra para decisiones médicas a menudo dependen de factores “de confusión”, como el tipo de equipo de rayos X utilizado¹⁵⁸ o la fuga de datos, lo que los haría potencialmente inútiles en la práctica, y estos factores de confusión son extremadamente difíciles de detectar sin un algoritmo interpretable, y tratar de explicar estas cajas negras a menudo no es suficiente. Hay una gran cantidad de personas que entrenan redes neuronales de caja negra cuando bastarían modelos interpretables que son mucho más fáciles de entrenar. Muy pocas personas intentan entrenar una red neuronal interpretable. Por otro lado, cuando se hace correctamente, los métodos de aprendizaje automático pueden tener grandes éxitos, como ahorro de recursos, reducción de tiempos y ampliación de las capacidades para investigar temas o resolver problemas en prácticamente cualquier sector de la industria, el gobierno y la sociedad en general.

La capacidad de detectar patrones de ML, en el caso de que los datos tengan un patrón verificado o verificable lo hace muy confiable y útil. Pero en el caso de que exista alguna falla en los datos de alguna variable, el modelo va a tomar el indicador que cree más importante y hará sus predicciones para el mundo real sin dicha precisión. El problema está en que los algoritmos de ML no tienen manera de

¹⁵⁶ Frank Pasquale, “Machines Judging Humans: The Promise and Perils of Formalizing Evaluative Criteria”, en *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*, Association for Computing Machinery, New York, 2020, p. 7. DOI:<https://doi.org/10.1145/3375627.3375839>. [20 de julio de 2022]

¹⁵⁷ J. Destin, *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters, octubre 9, 2018.

¹⁵⁸ J. Zech, M. Badgeley, M. Liu, A. Costa, J. Titano, and E. Oermann. *Variable generalization performance of a Deep Learning model to detect pneumonia in chest radiographs: A cross-sectional study*. PLoS Med., 15(e1002683), 2018.

distinguir si una variable está correlacionada con otra por “casualidad” o por un “error en el procesamiento de datos”, mientras que un humano sí puede distinguirlo simplemente con el mero uso de su experiencia en este tema o de su intuición. Los métodos bayesianos permiten perfeccionar esto, pero querer erradicar por completo la falla nos forzará a rediseñar desde cero los algoritmos de utilizados. ¹⁵⁹

Para finalizar esta sección debemos decir que las metodologías basadas en ML – DL están intentando revolucionar los servicios de salud; su uso permite a las máquinas aprender de los datos para hacer predicciones. Aquí, el ML – DL cambia el paradigma de la programación tradicional que se utiliza en los Sistemas Expertos Decisionales, en el que las instrucciones se debían de dar al programador y éste a la máquina.

Podemos decir que un Sistema de Apoyo a la Decisión Clínica impulsado por Inteligencia Artificial (SADC) aprovecha la capacidad de procesamiento de la máquina y el resultado se suma a la capacidad de razonamiento del médico. Su unión permite un diagnóstico y/o tratamiento con un grado de mayor precisión. La máquina y el médico están buscando patrones/síntomas, pero el médico no puede analizar todos los patrones/síntomas de todas las enfermedades que se generan ni puede revisar todos los estudios de todos los pacientes en un tiempo reducido. La máquina aquí tiene su utilidad.

En la siguiente tabla podemos apreciar algunos ejemplos de las principales técnicas de ML que son utilizadas para determinadas actividades de detección, análisis o diagnóstico médico:

¹⁵⁹ Algunos expertos sostienen que los modelos como *Shap* o *Lime* pueden decir con precisión qué variables son las más influyentes y esto, sumando a la experiencia del técnico del negocio le permitirá identificar las variables que se deben a la causalidad.

Técnicas de ML utilizadas para la predicción de enfermedades

Actividad	Técnica de ML
Detección de enfermedades cardíacas	• Clasificador Naive Bayes (NBC) • Máquinas de Vector Soporte (Support Vector Machine, SVM) • Red Bayesiana
Análisis de enfermedades diabéticas	• Algoritmo de clasificación Naive Bayes (NBC) • Árbol de decisiones • Árbol CART (Classification and Regression Trees)
Diagnóstico de trastorno tiroideo	• Máquina de Vector Soporte (Support Vector Machine, SVM) • Árbol de decisiones
Aprendizaje automático en cáncer de mama ¹⁶⁰	• Algoritmo J48 • Algoritmo de clasificación Naive Bayes (NBC) • Máquinas de Vector Soporte (Support Vector Machine, SVM) • Árbol de decisiones • Vecinos más cercanos (K Nearest Neighbours K-NN)
Diagnóstico de melanoma maligno ¹⁶¹	• Redes Neuronales Convolucionales Profundas

Fuente: El autor con base en: K. Shailaja, B. Seetharamulu and M. A. Jabbar, "Machine Learning in Healthcare: A Review," *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018, pp. 910-914, doi: 10.1109/ICECA.2018.8474918. [15 julio de 2022]

Un sistema de apoyo a la toma de decisiones para el diagnóstico o tratamiento de enfermedades basado en ML – DL ayuda al médico, a la organización y, principalmente, al paciente. También, se puede encargar del registro de imágenes y recuperación de información de expedientes, bases de datos y estudios clínicos.

En el caso del sector médico, por los datos que se utilizan, se ha hecho mucho énfasis en el aprendizaje automático interpretable, es decir, aquellos modelos que pueden proporcionar explicaciones sobre por qué se hacen ciertas predicciones o el cómo se llega a ellos. Los proveedores clínicos y otros responsables de la toma de decisiones en *healthcare* están poniendo el tema de la interpretabilidad como una prioridad para su implementación y utilización.

¹⁶⁰ Los datos son tomados de Wisconsin Dataset UCI, repositorio de *Machine Learning*: <https://archive.ics.uci.edu/ml/index.php> [20 de julio de 2022]

¹⁶¹ Ningrum DNA, Yuan SP, Kung WM, Wu CC, Tzeng IS, Huang CY, Li JYC, Wang YC, "Deep Learning Classifier with Patient's Metadata of Dermoscopic Images in Malignant Melanoma Detection". *J Multidiscip Healthc*. Número 14, 2021, pp. 877-885. Consúltase en: <https://doi.org/10.2147/JMDH.S306284>. [20 de julio de 2022]

A medida que las aplicaciones de aprendizaje automático se están integrando, la necesidad de explicación de la predicción crece.

Una barrera importante de la IA en la salud, especialmente en los modelos de Redes Neuronales Profundas, es su característica de caja negra. Es decir, el funcionamiento interno del algoritmo que incluye la justificación de cómo se generan las predicciones y cómo se consideran las diferentes características clínicas en el modelo permanece opaco para los profesionales médicos. Esto de pronto se convierte en obstáculo para su uso ante una comprensión incompleta y la falta de confianza por los resultados obtenidos. El Aprendizaje automático interactivo¹⁶² puede ser la opción para el problema de la caja negra. A menudo se basa en técnicas avanzadas de visualización e interacción, lo que permite a los usuarios monitorear, interpretar e intervenir el proceso de aprendizaje automático para que puedan establecer confianza en los resultados del algoritmo.¹⁶³ En una línea similar, un grupo de investigadores de Google creó recientemente un sistema de visualización experimental para ayudar a los médicos a comprender las interpretaciones de la IA de los datos de imágenes de cáncer de mama.¹⁶⁴

Otro aspecto relevante de los sistemas de apoyo a las decisiones médicas que utilizan ML está en el sentimiento del médico y su percepción respecto del posible desplazamiento por la IA. Estudios recientes documentaron las preocupaciones de los profesionales médicos sobre la posibilidad de que la IA reemplace sus trabajos a medida que la IA se vuelve más poderosa y toma más

¹⁶² Jerry Alan Fails y Dan R Olsen Jr., “Interactive Machine Learning” en *Proceedings of the 8th international conference on Intelligent user interfaces*. ACM, 2003, pp. 39-45.

¹⁶³ Consúltese: Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. “Trust in AutoML: exploring information needs for establishing trust in automated Machine Learning systems” en *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020, 297-307. Shweta Narkar, Yunfeng Zhang, Vera Liao, Dakuo Wang, and Justin Weisz. “Model LineUpper: Supporting Interactive Model Comparison at Multiple Levels for AutoML” en *Proceedings of the ACM IUI 2021*, 2021. Daniel Karl I Weidele, Justin D Weisz, Erick Oduor, Michael Muller, Josh Andres, Alexander Gray, and Dakuo Wang. “AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates” en *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020, 308-312.

¹⁶⁴ Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, *et al.*, “Human-centered tools for coping with imperfect algorithms during medical decision-making”, en *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 4.

iniciativas que los propios médicos¹⁶⁵. Esto está relacionado con el concepto de autonomía de la voluntad, libre desarrollo de la personalidad y la toma de decisiones, así como la responsabilidad profesional y Ética en el desempeño de sus actividades. Los médicos prefieren asumir las consecuencias de sus actos y ser responsables directos de sus decisiones antes que un tercero decida (sea humano o máquina)¹⁶⁶. Sobre todo, en países o lugares no muy familiarizados con la tecnología hay una percepción de amenaza a la autonomía profesional y el grado en que los médicos sean responsables por el uso de un sistema que limita el control de las decisiones¹⁶⁷.

Para llevar estos conceptos a la práctica, diversos investigadores hicieron un trabajo de campo con 22 médicos en 6 clínicas rurales en China y reportaron las diversas tensiones entre el diseño de un sistema de apoyo a las decisiones médicas (*Brilliant AI Doctor*) y su relación con el contexto clínico rural, el contexto local, el flujo de trabajo, así como las limitaciones técnicas y barreras de uso (cuestiones relacionadas con la transparencia y confiabilidad). En general, los resultados fueron favorables al uso de un asistente de AI para los médicos en colaboración con humanos.¹⁶⁸

Los algoritmos de ML pueden ayudar a los pacientes y a sus médicos para la toma de decisiones sobre el curso de la atención tanto en situaciones paliativas

¹⁶⁵ Bo Gong, James P Nugent, William Guest, William Parker, Paul J Chang, Faisal Khosa, and Savvas Nicolaou, "Influence of Artificial Intelligence on Canadian Medical Students Preference for Radiology Specialty: ANational Survey Study" en *Academic radiology*, número 26, 4, 2019, pp. 566–577.

¹⁶⁶ Ten Have, Henk, "Re-evaluating professional autonomy in health care", *Theoretical medicine and bioethics* número 21, vol. 5, 2020, pp. 503–513.

¹⁶⁷ Consúltese: Friedberg, Mark W, Chen, Peggy G., et. al., "Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy", *Rand health quarterly*, número 3, 4, 2014. Sambasivan, Murali, Esmailzadeh, Murali et. al., "Intention to adopt clinical decision support systems in a developing country: effect of physician's perceived professional autonomy, involvement and belief: a cross-sectional study". *BMC medical informatics and decision making*, número 12, 1, 2012, p. 142.

¹⁶⁸ Dakuo, Wang, Liuping, Wang, et. al., "Brilliant AI Doctor" in Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment. In Proceedings of the 2021" en *Conference on Human Factors in Computing Systems (CHI '21)*, Association for Computing Machinery, article 697, New York, pp. 1-18. DOI: <https://doi.org/10.1145/3411764.3445432>

como en la recuperación misma. Un equipo de Google, en colaboración con tres centros médicos académicos, utilizó información de más de 216,000 hospitalizaciones de 114,000 pacientes y casi 47 mil millones de puntos de datos para predecir si un paciente muriera, la duración de la estadía, la readmisión hospitalaria inesperada y los diagnósticos finales de alta. Todos se predijeron con un rango de precisión que fue bueno y bastante consistente entre los hospitales que se estudiaron.¹⁶⁹

Un grupo alemán utilizó el aprendizaje profundo en más de 44.000 pacientes para predecir la muerte hospitalaria, la insuficiencia renal y las complicaciones hemorrágicas después de la cirugía con una precisión notable. La IA también se ha utilizado para predecir si un paciente sobrevivirá después de un trasplante de corazón y para facilitar la combinación de registros de salud electrónicos y datos de secuencia. Los algoritmos de ML pueden influir en la utilización de recursos para los sistemas de salud, como unidades de cuidados intensivos, reanimación o ventiladores.¹⁷⁰

Los hospitales son el rubro número uno de los costos de atención médica en los Estados Unidos. Si bien el personal es el factor más importante que determina sus costos, la necesidad de evitar las hospitalizaciones ya sea la primera o la readmisión, ha ocupado un lugar central en muchas iniciativas de ML - DL. La economía juega un papel importante aquí, ya que la readmisión dentro de los treinta días posteriores a la hospitalización puede no ser reembolsable. Existe preocupación, y de hecho cierto debate, sobre si tratar de restringir las hospitalizaciones tendría un efecto adverso en los resultados de los pacientes. Múltiples estudios han asumido el desafío de predecir si un paciente hospitalizado necesitará ser readmitido en el mes siguiente al alta del hospital. un hospital, en particular para encontrar características que no son capturadas por los médicos.¹⁷¹

¹⁶⁹ Eric, Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books. Edición de Kindle, 2019.

¹⁷⁰ *Idem.*

¹⁷¹ *Idem.*

Un estudio realizado por Mount Sinai en la ciudad de Nueva York utilizó registros médicos electrónicos, medicamentos, laboratorios, procedimientos y signos vitales, y demostró una precisión del 83 por ciento en una cohorte relativamente pequeña. Se utilizó un grupo mucho más grande de 300.000 pacientes para entrenar y validar el DNN de DeepR Analytics, que se comparó favorablemente con otros esfuerzos, incluidos DoctorAI¹⁷² y DeepCare. En particular, Intermountain Healthcare, University of Pittsburgh Medical Center y Sutter Health se encuentran entre los primeros en trabajar en la implementación de dichos algoritmos.¹⁷²

Entre los objetivos más atrevidos está predecir la enfermedad en pacientes sin ningún síntoma clásico. Un grupo con sede en la Universidad de Tsinghua en Beijing tomó datos de más de 18,000 HCE del mundo real para diagnosticar con precisión seis enfermedades comunes: hipertensión, diabetes, EPOC, arritmia, asma y gastritis. Usando únicamente dieciocho pruebas de laboratorio, ciertas condiciones, como enfermedad renal, podría predecirse con precisión mediante casi 300.000 pacientes seguidos durante ocho años. El grupo de Mount Sinai estudió las HCE de 1,3 millones de pacientes para predecir cinco enfermedades: diabetes, demencia, herpes zóster, anemia de células falciformes y trastorno por déficit de atención, con una precisión muy alta.¹⁷³

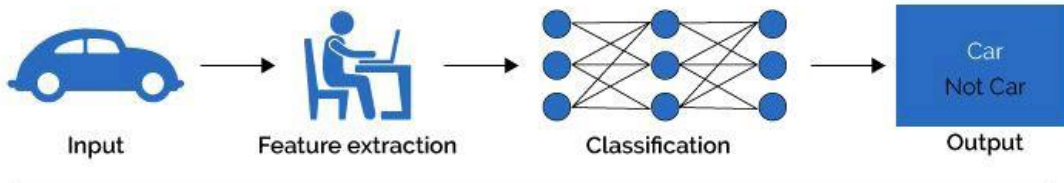
El ámbito predictivo es el que más resultados ha tenido por el uso de ML - DL. Lo cual no significa la derrota de los otros intentos, solamente será de tiempo para demostrar que la IA sirve en todo lo que se lo proponga el ser humano. Evidentemente, con las consecuencias naturales de su uso. El caso *Watson Health* es un ejemplo con muchos elementos para el análisis que se verá en una sección posterior. Por el momento piense qué preocupaciones le generaría si supiera que su vehículo funciona con esta tecnología y en el manual del usuario se le presenta la siguiente imagen:

¹⁷² *Idem.*

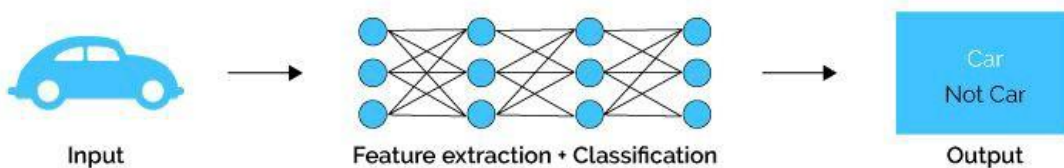
¹⁷³ *Idem.*

Machine Learning vs Deep Learning

Machine Learning



Deep Learning



Fuente: <https://lawtomated.com/a-i-technical-machine-vs-deep-learning/> [15 de julio de 2022]

Estudios demuestran que el ser humano, prefiere mantener el control de las máquinas en sus manos y con su voluntad, a pesar de saber que la máquina lo supera en precisión y eficiencia de movimientos. ¿Qué pensaría si en lugar de un vehículo fuera un robot el que haga la operación de extirpación de algún tumor cancerígeno? Bajo el mismo esquema de la imagen, pero ahora, ¿la máquina con ayuda de un sistema inteligente?

Capítulo II. Inteligencia Artificial basada en Machine Learning – Deep Learning. Problemas de funcionamiento y rendición de cuentas

Capítulo II. Inteligencia Artificial basada en Machine Learning – Deep Learning.

Problemas de funcionamiento y rendición de cuentas

De 2006, año en que Geoffrey Hinton relanzó las redes neuronales con su modelo de aprendizaje profundo, a la fecha de la conclusión de nuestra investigación (julio de 2022), la evolución de la IA basada en Machine Learning – Deep Learning ha cambiado más que durante todo el siglo XX. Sin embargo, como toda obra humana, esta tecnología no es perfecta y tiene limitantes que, desde nuestro punto de vista, generan la problemática de afectación a Derechos Humanos. Por dar un ejemplo: Un estudio hecho por *The Royal Society*¹⁷⁴, señala que existen cuatro limitantes que generan problemas por el uso de la IA basada en Machine Learning – Deep Learning:

- a) Datos: En particular, aquellos que son utilizados para predicciones. Hay algoritmos que trabajan con datos y otros sin datos; y ambos pueden producir información de mayor calidad y en menor tiempo o causar problemas que afecten Derechos Humanos al generar sesgos por el uso de esta tecnología.
- b) Analogía: Los seres humanos son buenos para transferir ideas por analogía de un dominio a otro o de un problema a otro. Esto sigue siendo un desafío para las computadoras, incluso con las últimas técnicas de aprendizaje automático de pronto no son tan precisas en cuestiones hechas por el hombre con base en razonamiento abductivo o heurístico.
- c) Interpretación del actuar humano: Lograr conocer la verdadera intención de los humanos es muy complejo, requiere una comprensión sofisticada de la mente que en ocasiones ni los propios seres humanos logran entender en el trato cotidiano.
- d) Contexto: Es difícil desarrollar sistemas con comprensión del contexto del problema o lo que los seres humanos conocemos como sentido común, intuición o razonamiento abductivo. Si bien se está trabajando en este tema,

¹⁷⁴ Cfr. *Machine Learning: the power and promise of computers that learn by example*, Londres, 2017.

sigue en desarrollo para lograr que los sistemas de aprendizaje automático logren alcanzar este tipo de razonamiento.

Respecto a este último inciso d) y su relación con el inciso c), Gerd Gigerenzer, comenta que la psicología social en algún momento buscó demostrar que las intuiciones fallaban, que eran irracionales. La hipótesis surge con Daniel Kahneman y Amos Tversky (Teoría de las Perspectivas), quienes afirman que los “individuos toman decisiones, en entornos de incertidumbre, que se apartan de los principios básicos de la Teoría de la Economía Estándar. A este tipo de decisiones lo denominaron atajos heurísticos, es decir, aquellos pensamientos que siempre evitarán una pérdida. Una persona prefiere no perder 50 mil pesos antes que ganar 50 mil pesos.”¹⁷⁵ Sin embargo, ahora tenemos nuevos resultados, que están alejándose de esa idea, y se está tratando de demostrar que la intuición puede ser mejor que los modelos de elección racional, o la regresión múltiple o los programas estadísticos más complejos.

Esto no se sabía en el siglo XX, se creía que la toma de decisiones debía ser consciente y guiarse por las leyes de la lógica, que era recomendable elaborar listas con los pros y contras para tomar la mejor decisión. Ahora se sabe que las decisiones intuitivas son eficaces, a veces mucho más que una elección racional. Buena parte de las decisiones que tomamos son intuitivas y alejadas de la lógica, rápidas, aparecen y nos conducen a la respuesta acertada. ML – DL ya están haciendo predicciones intuitivas y en algunos temas están utilizando más percepción que la lógica tradicional. Es un tema que está estudiándose y que seguramente tendrá más avances en poco tiempo.

Por el momento, de la revisión histórica hecha debemos decir que, Gigerenzer y su equipo de investigadores demostraron que es mejor tomar una

¹⁷⁵ Daniel Kahneman, titular de la cátedra Eugene Higgins de Psicología en la Universidad de Princeton en Nueva York y profesor de Relaciones Públicas en la Woodrow Wilson School de Princeton, recibió el Premio Nobel de Economía por su trabajo con Amos Tversky sobre los sesgos cognitivos (2002). Sus ideas han tenido un profundo impacto en campos tan diversos como la economía, la medicina o la política. Cfr. Gerd, Gigerenzer, *Decisiones intuitivas*, Ariel. Edición de Kindle, 2018. La intuición no es racional [Entrevista a Gigerenzer] en el sitio: https://www.youtube.com/watch?v=6fjMd63ULf8&feature=emb_logo y también puede consultarse: El País Semanal: ¿Podemos fiarnos de la Intuición?, Jenny Moix Queraltó consúltase en: https://elpais.com/diario/2010/11/14/eps/1289719613_850215.html.

decisión por intuición que hacer toda una operación estadística complicada. Aquí se aplica lo que se denomina heurística de conocimiento o cognición heurística (escoge lo que conozcas; la intuición se basa en principios sencillos que ignoran información y basarse solo en lo que conoces es un principio heurístico). La ignorancia parcial puede ser útil en algunas ocasiones.¹⁷⁶

A continuación, explicaremos cómo influye el uso de datos en la toma de decisiones y la generación de sesgos por los datos con los que se trabaja y que proporcionan las personas; por el propio programador del sistema, que deja (inconscientemente) sus sesgos humanos; o, por la interacción entre el usuario y el sistema. Datos y sesgos sumados al factor humano generan los principales problemas que tiene la IA basada en Machine Learning – Deep Learning en su enseñanza, creación y aplicación; ahí encontramos las diferentes propuestas de regulación que existen y que están en proceso de aprobación por organismos nacionales y supranacionales.

En esta sección se expondrán: A) Problemas de funcionamiento de esta tecnología, en particular, aquellos relacionados con 1) los datos y sesgos; y, 2) aquellos que surgen de la opacidad y que nos llevan a la interpretabilidad y a la explicabilidad algorítmica. También, expondremos la B) Rendición de cuentas algorítmica que existe actualmente, tanto en normas relacionadas con Transparencia y Privacidad 1) como en aquellas relacionadas con la auditabilidad algorítmica (2).¹⁷⁷

¹⁷⁶ Cfr. Gerd, Gigerenzer, *Decisiones instintivas*, Ariel. Edición de Kindle, 2018. La intuición no es racional [Entrevista a Gigerenzer] en el sitio: https://www.youtube.com/watch?v=6fjMd63ULf8&feature=emb_logo y también puede consultarse El País Semanal: ¿Podemos fiarnos de la Intuición? De Jenny Moix Queraltó https://elpais.com/diario/2010/11/14/eps/1289719613_850215.html

¹⁷⁷ Aquí hacemos una advertencia, esta investigación está acotada en su temporalidad a enero de 2022. La experiencia nos dice que estos temas se mueven y cambian vertiginosamente, lo cual debe tener presente el lector para fijar su crítica o análisis en el período comentado (verano de 1956 a verano de 2021). Dar seguimiento y actualizar estos temas a posteriores avances supera las posibilidades del autor.

1) *Problemas de funcionamiento*

La problemática en el uso de la IA basada en Machine Learning – Deep Learning será acotada, para efectos de este trabajo, en el uso que se le da a los datos y los problemas que generan los sesgos. ¿Qué debemos entender por datos? Una definición gramatical de Dato y de algunas voces que utilizamos: 1. m. Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho. A este problema le faltan datos numéricos. 2. m. Documento, testimonio, fundamento. 3. m. Inform. Información dispuesta de manera adecuada para su tratamiento por una computadora. 1. m. Archivo de datos referidos a una determinada materia, que puede ser utilizado por diversos usuarios. Por base de datos 1. f. Inform. Conjunto de datos organizado de tal modo que permita obtener con rapidez diversos tipos de información. Por procesamiento de datos 1. m. Inform. Aplicación sistemática de una serie de operaciones sobre un conjunto de datos, generalmente por medio de máquinas, para explotar la información que estos datos representan. Por protección de datos. 1. f. Sistema legal que garantiza la confidencialidad de los datos personales en poder de las Administraciones públicas u otras organizaciones.¹⁷⁸

Visto el tema desde el ámbito de la IA y más concretamente desde la Lógica de la programación, existen varios conceptos que debemos conocer:

I. Atributo. - aquella característica que identifica a un ente informático;

II. Ente informático. - es todo aquello que puede describirse basándose en sus características. Sus características están reflejadas en datos:

III. Dato. - es el atributo codificado en términos que sean entendibles dentro de un sistema informático (programa de software), manejables y comparables (quizás hasta absolutos, en gran medida);

¹⁷⁸ Cfr. *Diccionario de la Lengua Española*, edición digital en: <https://dle.rae.es/dato?m=form> [20 de julio de 2022]

IV. Campo. - es el nombre que se le coloca a un dato para identificar el atributo;

V. Registro. - es el conjunto de campos en el que se encuentra consignado un dato;

VI. Archivo. - se integra por un conjunto de registros que tienen la misma escritura y que se puede manejar como una sola unidad; y

VII. Base de Datos. - La información almacenada en varios archivos interrelacionados.¹⁷⁹

Por otro lado, el concepto información es un conjunto de datos suficientemente organizados y entendibles; y, la “informática” es la ciencia que estudia, aplica y optimiza el tratamiento eficiente de la información.¹⁸⁰ La RAE define el concepto información como el “conjunto de conocimientos científicos y técnicas que hacen posible el tratamiento automático de la información por medio de computadoras”.¹⁸¹

Existe también un ingrediente importante denominado Lenguaje de Programación (*Programming Language*), es decir, aquel conjunto de símbolos que se utiliza para definir una secuencia de instrucciones, a fin de que puedan ser procesadas en una máquina. La definición de un lenguaje de programación conlleva, *per se*, al riesgo de cometer errores (o sesgos) que pudieran considerarse como un factor que afecta los resultados de predicción:

- Errores de concepción: Cuando el programador cree conocer bien el objetivo, sin embargo, no es así. Finaliza el programa con algo que no resuelve el problema.

¹⁷⁹ Trejos Buriticá, Omar, *Lógica de la programación*, Bogotá, 2017, pp. 95 a 103. “En su forma más básica, un conjunto de datos se organiza en una matriz $n \times m$, donde n es el número de instancias (filas) y m es el número de características (columnas). Kelleher, John D. *Deep Learning* (serie The MIT Press Essential Knowledge) (p. 252). Prensa del MIT. Edición de Kindle.

¹⁸⁰ Ibidem., p. 23.

¹⁸¹ Informática o informático, voces consultadas en: Real Academia Española, *Diccionario de la lengua española*, 23ª edición, [versión 23.5 en línea]. Consúltese en: <https://dle.rae.es/inform%C3%A1tico> [Fecha de consulta 20 de julio de 2022]

- Errores lógicos: Cuando los algoritmos no están bien contruidos y presentan fallas en su orden o en su inicio y fin.
- Errores de procedimiento: Cuando se cree que el algoritmo está bien, se hace la prueba de escritorio para verificarlo y se cree que funciona bien.
- Errores de sintaxis: Cuando realizamos la programación podemos tener errores en la transcripción
- Errores semánticos: Cuando la sintaxis del código está bien articulada, pero el significado no es el que se buscaba. ¹⁸²

A continuación, explicaremos los datos y sesgos y, en segundo lugar, la opacidad como un problema de la IA basada en Machine Learning – Deep Learning que nos lleva al entendimiento de los conceptos interpretabilidad y explicabilidad algorítmica.

A) Datos y sesgos

Actualmente, los datos que generamos los seres humanos surgen de diversas fuentes, desde nuestro equipo de cómputo, teléfono móvil, tableta hasta en las tiendas de comercio minorista, escuelas, hospitales, gimnasio, bancos, servicios públicos, seguros; y en prácticamente cualquier actividad que esté vinculada a un equipo electrónico. Esto traerá como consecuencia un enorme volumen de datos para los individuos, organizaciones y gobiernos.

El Libro Blanco sobre la inteligencia artificial -un enfoque europeo orientado a la excelencia y la confianza- habla de los datos y su uso en sistemas de IA. El Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial explica lo siguiente: “...los sistemas de inteligencia artificial (IA) son programas informáticos (y posiblemente también equipos informáticos) diseñados por seres humanos que [...] actúan en la dimensión física o digital mediante la percepción de su entorno mediante la adquisición de datos, la interpretación de los datos estructurados, el razonamiento sobre el conocimiento o el tratamiento de la información, fruto de estos datos y las mejores acciones que se llevarán a cabo para alcanzar el objetivo

¹⁸² Trejos Buriticá, Omar, *Lógica de la programación*, Bogotá, 2017, pp. 95 a 103.

fijado.”¹⁸³ Esta definición incorpora los tres tipos de datos que existen o que son utilizados por los científicos de datos:

DATOS ESTRUCTURADOS La mayoría de las fuentes de datos tradicionales son datos estructurados, datos con formato o esquema fijo que poseen campos fijos. En estas fuentes, los datos vienen en un formato bien definido que se especifica en detalle, y que conforma las bases de datos relacionales. Son los datos de las bases de datos relacionales, las hojas de cálculo y los archivos, fundamentalmente. Los datos estructurados se componen de piezas de información que se conocen de antemano, vienen en un formato especificado, y se producen en un orden especificado. Estos formatos facilitan el trabajo con dichos datos. Formatos típicos son: fecha de nacimiento (DD, MM, AA); documento nacional de identidad o pasaporte (por ejemplo, 8 dígitos y una letra); número de la cuenta corriente en un banco (20 dígitos), etcétera.

DATOS SEMIESTRUCTURADOS Los datos semiestructurados tienen un flujo lógico y un formato que puede ser definido, pero no es fácil su comprensión por el usuario. Datos que no tienen formatos fijos, pero contienen etiquetas y otros marcadores que permiten separar los elementos de datos. La lectura de datos semiestructurados requiere el uso de reglas complejas que determinan cómo proceder después de la lectura de cada pieza de información. Un ejemplo típico de datos semiestructurados son los registros Web logs de las conexiones a Internet. Un Web log se compone de diferentes piezas de información, cada una de las cuales sirve para un propósito específico. Ejemplos típicos son el texto de etiquetas de lenguajes XML y HTML. Datos que no tienen formatos fijos, pero contienen etiquetas y otros marcadores que permiten separar los elementos de datos. Ejemplos típicos son el texto de etiquetas de XML y HTML.

DATOS NO ESTRUCTURADOS Los datos no estructurados son datos sin tipos predefinidos. Se almacenan como “documentos” u “objetos” sin estructura uniforme, y se tiene poco o ningún control sobre ellos. Datos de texto, video, audio, fotografía son datos no estructurados. Por ejemplo, las imágenes se clasifican por su resolución en píxeles. Datos que no tienen campos fijos; ejemplos típicos son: audio, video, fotografías, documentos impresos, cartas, hojas electrónicas, imágenes digitales, formularios especiales, mensajes de correo electrónico y de texto, formatos de texto libre como correos electrónicos, mensajes instantáneos SMS, artículos, libros, mensajes de mensajería instantánea tipo WhatsApp, Une, Joyn, Viber, Une, WeChat, Spotbros. Al menos, el 80% de la información de las organizaciones no reside en las bases de datos relacionales o archivos de datos, sino que se encuentran esparcidos a lo largo y ancho de la organización; todos estos datos se conocen como datos no estructurados. Sin duda, los datos más difíciles de dominar por los analistas son los datos no estructurados, pero su continuo

¹⁸³ Comisión Europea, *Libro Blanco... op. cit.*, p. 20. Esta definición tiene un antecedente previo en el documento de trabajo “Una Definición de la Inteligencia Artificial: Principales capacidades y disciplinas científicas, del Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial que a su vez toma la definición de la Comunicación de la Comisión, de 25 de abril de 2018, titulada “Inteligencia Artificial para Europa” (COM(2018)0237): “...se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción -con cierto grado de autonomía- con el fin de alcanzar objetivos específicos...”. Comunicación de la Comisión Europea sobre la IA al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia Artificial para Europa, Bruselas, 25.4.2018 COM(2018) 237 final. La segunda definición, que podemos leer aquí, incluye *Machine Learning* y *Deep Learning* [El énfasis es propio]

crecimiento ha provocado el nacimiento de herramientas para su manipulación como es el caso de MapReduce, Hadoop o bases de datos NoSQL...¹⁸⁴

También se puede hacer una clasificación de los datos de la *Web* como la fuente de mayor riqueza que existe en el mundo. El internet de las cosas, la geolocalización satelital y el uso de biométricos va a generar un incremento adicional. Una forma de clasificar esto es la siguiente:

- 1.- Web and Social Media:** Incluye contenido web e información que es obtenida de las redes sociales como Facebook, Twitter, LinkedIn, etc, blogs.
- 2.- Machine-to-Machine (M2M):** M2M se refiere a las tecnologías que permiten conectarse a otros dispositivos. M2M utiliza dispositivos como sensores o medidores que capturan algún evento en particular (velocidad, temperatura, presión, variables meteorológicas, variables químicas como la salinidad, etc.) los cuales transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa.
- 3.- Big Transaction Data:** Incluye registros de facturación, en telecomunicaciones registros detallados de las llamadas (CDR), etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados.
- 4.- Biometrics:** Información biométrica en la que se incluyen huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc. En el área de seguridad e inteligencia, los datos biométricos han sido información importante para las agencias de investigación.
- 5.- Human Generated:** Las personas generamos diversas cantidades de datos como la información que guarda un call center al establecer una llamada telefónica, notas de voz, correos electrónicos, documentos electrónicos, estudios médicos, etc.¹⁸⁵

Y el *IEEE SA Standards Board*, al emitir su “Recomendación para evaluar el impacto de los sistemas autónomos e inteligentes en el bienestar humano” (*human well-being*) proporciona consejos para el uso de los datos a creadores, diseñadores, desarrolladores, ingenieros, programadores, entre otros. Estos conocimientos sobre el impacto que debe tenerse en cuenta durante todo el ciclo de vida de estos sistemas y hace un modelo experimental para medir el bienestar humano se basan en el principio de la privacidad (que se verá más adelante) y en el respeto a la propiedad de los datos de conformidad con las leyes aplicables. En su sección

¹⁸⁴ Luis Joyanes, *Big Data - Análisis de grandes volúmenes de datos en organizaciones*, Alfaomega. Edición de Kindle, pp. 3-5.

¹⁸⁵ Ricardo, Barranco Fragoso, “¿Qué es el Big Data? Todos formamos parte de ese gran crecimiento de datos”, IBM Corporation, 2012, pp. 3-4. <https://developer.ibm.com/es/technologies/data-science/articles/que-es-big-data/>

“definiciones” aporta conceptos como: “Datos de referencia: datos recopilados al comienzo de un proceso y uso para compararlos con los datos recopilados posteriormente; datos primarios: datos que se recopilan directamente de un sujeto y, datos secundarios: datos que ya se han recopilado de diversas fuentes, como censos gubernamentales u oficinas de estadística, organizaciones sin fines de lucro, instituciones académicas y consultorías”.¹⁸⁶

El WIA o *Well-Being Impact Assessment* es el proceso mediante el cual uno identifica los impactos potenciales en el bienestar a través de la participación de las partes interesadas, mide los impactos a través de la recopilación y el análisis de datos, y modifica un sistema de IA con base en los conocimientos derivados de la participación de las partes interesadas y el análisis de datos. Este proceso es iterativo, ya que debe repetirse a lo largo del ciclo de vida del sistema y se compone de cinco actividades: a) Análisis interno y participación de las partes interesadas / usuarios; b) Creación del tablero de indicadores de bienestar; c) Creación de un proceso de recopilación de datos y la recopilación de datos; d) Análisis de datos y uso de dichos datos para mejorar el sistema; y e) Iteración.¹⁸⁷

Visto el tema de los datos desde el enfoque del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos, RGPD) debemos tomar en cuenta que existen datos personales y el procedimiento de protección que obliga a los sujetos que los utilizan a dar un buen tratamiento. La protección de los datos personales de las personas físicas es un derecho fundamental contenido en el artículo 8, apartado 1, de la Carta de los Derechos Fundamentales de la Unión

¹⁸⁶ Cfr. Standards Committee of the IEEE Systems, Man, and Cybernetics Society, “IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being”, *IEEE SA Standards Board*, 2020. doi: 10.1109/IEEESTD.2020.9084219 [La traducción es propia] <https://standards.ieee.org/content/ieee-standards/en/standard/7010-2020.html>

¹⁸⁷ *Idem*.

Europea y el artículo 16, apartado 1, del Tratado de Funcionamiento de la Unión Europea¹⁸⁸.

El tratamiento de datos personales de personas físicas debe de regular la recolección, uso, consulta y la medida en que dichos datos son o serán tratados. Existe en este tema el llamado principio de transparencia que exige que toda información y comunicación relativa al tratamiento de dichos datos sea accesible y fácil de entender, y que se utilice un lenguaje sencillo y claro. Aunque explicaremos este principio en un apartado posterior, debemos dejar claro desde ahora que se refiere a la información de los interesados sobre la identidad del responsable del tratamiento y los fines de este y a la información añadida para garantizar un tratamiento transparente con respecto a las personas físicas afectadas y a su derecho a obtener confirmación y comunicación de los datos personales que les conciernan.

Otra característica es que los datos personales deben ser “adecuados, pertinentes y limitados a lo necesario para los fines para los que sean tratados. Ello requiere, en particular, garantizar que se limite a un mínimo estricto su plazo de conservación. Los datos personales solo deben tratarse si la finalidad del tratamiento no pudiera lograrse razonablemente por otros medios”¹⁸⁹. Para garantizar que los datos personales no se conservan más tiempo del necesario, el responsable del tratamiento ha de establecer plazos para su supresión o revisión periódica. Deben tomarse todas las medidas razonables para garantizar que se rectifiquen o supriman los datos personales que sean inexactos. Los datos personales deben tratarse de un modo que garantice una seguridad y confidencialidad adecuadas de los datos personales. Para que el tratamiento sea

¹⁸⁸ Desde el año 2009, el artículo 6, segundo párrafo de la Constitución Política de los Estados Unidos Mexicanos contempla el mismo derecho de protección de datos: “...Toda persona tiene derecho a la protección de sus datos personales, al acceso, rectificación y cancelación de los mismos, así como a manifestar su oposición, en los términos que fije la ley, la cual establecerá los supuestos de excepción a los principios que rijan el tratamiento de datos, por razones de seguridad nacional, disposiciones de orden público, seguridad y salud públicas o para proteger los derechos de terceros.”

¹⁸⁹ Ver artículo 5 del RGPD.

lícito, los datos personales deben ser tratados con el consentimiento del interesado o sobre alguna otra base legítima establecida conforme a Derecho.

El RGPD, en su artículo 4, nos proporciona diversas definiciones que tienen relación con el uso de sistemas de IA:

“Datos personales: toda información sobre una persona física identificada o identificable (el interesado); ... mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona.

Tratamiento: cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción. Los datos que están siendo procesados por un sistema de ML deben estar sujetos a un tratamiento de la información privada (datos personales).

Datos biométricos: aquellos datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos.

Datos relativos a la salud de las personas: aquellos datos personales relativos a la salud física o mental de una persona física, incluida la prestación de servicios de atención sanitaria, que revelen información sobre su estado de salud”.¹⁹⁰

Los datos son el combustible con el cual funcionan los sistemas de IA basada en Machine Learning – Deep Learning. Los datos representan la posibilidad de decidir correctamente, equitativamente, eficaz y eficientemente, con respeto a la moral o siguiendo la norma jurídica. Cada ser humano procesa los datos del mundo de manera subjetiva, nada nos garantiza que lo que uno piensa como emisor del mensaje, llegue perfectamente o al cien por ciento igual al momento que lo codificamos e ingresa al receptor del mensaje. En consecuencia, los datos que utilizan los sistemas IA basada en Machine Learning – Deep Learning van de origen o, de manera natural y constante, con alguna desviación, tendencia, perjuicio o sesgos. Pero ¿qué es un sesgo?

¹⁹⁰ En todos estos temas puede consultarse el documento emitido por el *Commissioner for Human Rights, Unboxing Artificial Intelligence: 10 steps to protect Human Rights. Recommendation, Council of Europe*, 2019.

Un sesgo, según su definición gramatical es un “error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras”.¹⁹¹ Para los efectos de esta investigación, para definir el concepto de “sesgo algorítmico” debemos primero entender que la “justicia algorítmica” es una rama nueva del conocimiento jurídico que explora las preocupaciones técnicas, legales y éticas que existen en la toma de decisiones con base en algoritmos, en particular cuestiones de sesgos algorítmicos, que pueden afectar Derechos Humanos.¹⁹²

Consideramos como “sesgo algorítmico”, para los fines de este trabajo, aquel funcionamiento del sistema de IA basada en Machine Learning – Deep Learning que no genera los resultados esperados afectando con ello, directa o indirectamente, a personas o a ciertos subgrupos sociales, abarcando disparidades junto con atributos que son categorías protegidas en los tratados y convenciones internacionales de Derechos Humanos, como la clase socioeconómica o el nivel educativo, el sexo, la raza, la orientación sexual, la edad, alguna discapacidad y otros atributos relacionados con la dignidad de la persona y sus Derechos Humanos (la vida, la libertad, la propiedad, etcétera).¹⁹³

La “toma de decisiones algorítmica” se refiere a decisiones tomadas utilizando un sistema IA basada en Machine Learning – Deep Learning. Estas decisiones pueden automatizarse completamente o utilizarse para informar a los tomadores de decisiones humanos. Un algoritmo de filtrado de currículos sería un ejemplo del primero, porque automatiza las decisiones iniciales de selección de empleo; Un algoritmo de evaluación del riesgo de reincidencia sería un ejemplo de lo segundo, porque proporciona recomendaciones a los jueces para las decisiones de detención y libertad condicional.¹⁹⁴

¹⁹¹ Cfr. Real Academia Española, *Diccionario de la lengua española*, 23ª edición, [versión 23.5 en línea]. Consúltase en: <https://dle.rae.es/sesgo> [Fecha de consulta 22 de 20 de julio de 2022]

¹⁹² La construcción es propia, se hizo con base en algunas aportaciones de Alice Xiang, en su artículo “Reconciling Legal and Technical Approaches to Algorithmic Bias”, publicado el 4 de enero de 2021, en el volumen 88, de la *Tennessee Law Review* número 649 año 2021, pp. 10-11. Consúltase en SSRN: <https://ssrn.com/abstract=3650635> [20 de julio de 2022]

¹⁹³ La definición es propia tomando algunos términos de Alice, Xiang, Reconciling Legal and Technical Approaches to Algorithmic Bias (4 de enero de 2021). 88 *Tennessee Law Review* 649 (2021), pp. 10-11. Consúltase en SSRN: <https://ssrn.com/abstract=3650635> [20 de julio de 2022]

¹⁹⁴ *Idem*.

Podemos decir que existen sesgos generados por las personas, sesgos generados por las máquinas/sistemas y sesgos de la sociedad. En el primer caso, la frecuencia con la que las personas escriben/hablan sobre acciones, resultados o propiedades de algo no es un reflejo de las frecuencias del mundo real o el grado en el que una propiedad es característica de una clase de individuos. Es decir, la información que registramos sobre el mundo tiene un grado de sesgo determinado por el observador. En el segundo caso, se puede tomar como base la definición estadística de sesgo y/o del tipo de tecnología que se está utilizando. Así, habrá tantos sesgos como actividades en la que se pueda utilizar el ML y el sesgo así consistiría en la diferencia entre el valor estimado y el valor real. Finalmente, el sesgo social es una inclinación o prejuicio a favor o en contra de una persona o grupo, especialmente de una manera que se considera injusta en la sociedad o parte de ella.¹⁹⁵

Los sesgos se pueden producir, de manera constante o inconsciente, en todos los seres humanos, es algo que puede evitarse, no desaparece por completo (salvo que no se realice la actividad), mucho menos cuando se hacen actividades de probabilidad o estadística. Aunque sea exagerado, estarán siempre presentes y estarán dentro de las máquinas ya que el ser humano las programa, la crea, las modifica, las utiliza. Existen diferentes casos que se han venido analizando que derivan de dilemas que van surgiendo de estas reflexiones, como, por ejemplo, el *Moral Machine*¹⁹⁶ y la posibilidad de que un coche autónomo tenga la opción de tomar decisiones al momento de ser conducido.

Lo interesante también está en el propio sistema y su capacidad para la detección de sesgos y corregirlos para evitar que en la práctica volvieran a ocurrir o que otro sistema se utilice para estos fines. Se parte de una base, que el ser humano a veces cambia o no cambia a voluntad de otro ser humano. Debemos reconocer esto para que las máquinas nos faciliten la vida con sus capacidades, pero alguien

¹⁹⁵ Para una explicación más detallada del tema de los sesgos consúltese: Drew Roselli, Jeanna Matthews, and Nisha Talagala, "Managing Bias in AI. In Companion Proceedings" en *The 2019 World Wide Web Conference (WWW '19)*, Association for Computing Machinery, New York, 2019, pp. 539–544. DOI:<https://doi.org/10.1145/3308560.3317590> [20 de julio de 2022]

¹⁹⁶ <https://www.moralmachine.net/hl/es> [20 de julio de 2022]

debe cuidar que no tengan afecten estos sesgos, diluyendo la responsabilidad humana. Los límites los encontramos en lo social, legal, técnico, moral y en la forma en cómo logramos transparentar la solución o quizás evaluar si se produce una responsabilidad algorítmica. Todo esto lo analizaremos en un apartado próximo, por el momento debemos reconocer y tener presente que la tecnología de aprendizaje automático es por naturaleza sesgada.

Los sistemas de IA basada en Machine Learning – Deep Learning producen sesgos. Es decir, tienen datos de entrada que a su vez producen respuestas de salida con 1) información del ser humano, 2) de las bases de datos o 3) de su entorno. Existen opciones para modificar directamente los datos que alimentamos a través de técnicas como el incremento de datos para reducir sesgos o el ajuste/cambio de indicadores o de los parámetros de muestreo. Un ejemplo conocido es el tema de reconocimiento facial de Microsoft Surface. A este sistema le resultaba difícil entender y reconocer los rostros de las personas de ciertos grupos demográficos. La razón de esto era que el sistema carecía de suficientes datos representativos de los rostros de esas personas en el conjunto.

En el campo de la medicina, al momento de tomar decisiones médicas existen errores cognitivos que llevan a sesgos cuando se utilizan sistemas inteligentes basados en modelos matemáticos cuantitativos. Los médicos, en su actividad profesional, pocas ocasiones hacen cálculos formales -basados en Método Científico o en la estadística- para tomar decisiones en la atención de los pacientes. Su comprensión de los asuntos se basa en el conocimiento heurístico. La heurística implica conocer el mundo por la experiencia del ser humano y la intuición. Hay un conjunto de circunstancias, hechos, acontecimientos, (patrones sintomáticos) que el médico experimentado detecta en el paciente. Este tipo de razonamiento suele ser falible debido a errores cognitivos naturales que se encuentran presentes en todos los seres humanos y que generan sesgos en el diagnóstico y tratamiento. Los errores, por lo general surgen de ignorancia o del conocimiento inadecuado del tema o por una determinación/implementación inadecuada del conocimiento. Independientemente que se tocará en un apartado

posterior, es oportuno mostrar algunos de los principales tipos de errores cognitivos (sesgos cognitivos) del médico, que se han resumido en la siguiente tabla:

Tipos de errores cognitivos (sesgos cognitivos)

Tipo de error (sesgo)	Características
De disponibilidad	Ocurre cuando el médico toma la decisión con el primer caso que recuerda sin tomar en cuenta otros casos. Esto limita el diagnóstico al dato que recuerda ya sea por ser reciente o por ser de impacto.
De representación	Ocurre cuando el médico pone su atención a la presencia o ausencia de síntomas clásicos de una enfermedad sin tomar en consideración otros factores y su prevalencia en el paciente; ocurre también cuando el médico no admite que los resultados del diagnóstico (positivos) de una enfermedad rara son más propensos a que sean falsos positivos que positivos verdaderos.
Cierre prematuro	Ocurre cuando el médico hace un diagnóstico rápido por determinados síntomas (patrones) que dejan fuera otros diagnósticos posibles y detiene la búsqueda de opciones o datos para llegar a otro diagnóstico
De anclaje	Ocurren cuando el médico se empecina en mantener su diagnóstico con una impresión inicial del paciente sin reconocer que existen datos contradictorios o confusos. El médico descarta los datos confusos o contradictorios y decide no prestar atención a las alteraciones.
De confirmación	Ocurre a través de una selección de datos clínicos que respaldan la hipótesis deseada e ignoran aquellos con los que no concuerdan. El sesgo de confirmación se mezcla con un error de anclaje
De atribución	Ocurre cuando el médico toma su decisión basada en estereotipos negativos del paciente, lo cual puede llevarlo a ignorar o minimizar la posibilidad de una enfermedad diversa
Afectivo	Ocurre cuando el médico deja que los sentimientos personales (positivos o negativos) influyan sobre sus decisiones.

Fuente: El autor con base en: Douglas L. McGee, *Errores cognitivos en la toma de decisiones médicas*, DO, Emergency Medicine Residency Program, Albert Einstein Medical Center, 2018. Consúltase en: <https://www.msdmanuals.com/es/professional/temas-especiales/toma-de-decisiones-m%C3%A9dicas/errores-cognitivos-en-la-toma-de-decisiones-m%C3%A9dicas> [13 de julio de 2022]

En el mundo hay una tendencia a sustituir la racionalidad o la actividad humanas por una racionalidad o actividad tecnológica y esto ocurre en todos los campos del quehacer humano; la medicina no es la excepción, existen sesgos en las decisiones médicas y sesgos en los sistemas inteligentes que se utilizan para su apoyo. ¿Hasta qué punto esto sale del orden privado y entra al orden público? o mejor dicho ¿cómo entender el orden público si la dinámica tecnológica influye en las decisiones al grado de, en ocasiones, deja al ser humano fuera de cualquier intervención de su voluntad (directa o indirectamente)?

Los sistemas de IA basada en Machine Learning – Deep Learning están diseñados para mejorar la actividad de diagnóstico y tratamiento médicos, sin embargo, trabajan con datos imperfectos y con los sesgos humanos. Vista esta imperfección desde un enfoque jurídico: ¿En dónde queda la responsabilidad? En el diseñador, el científico de datos, programador, el sistema algorítmico, el propio dueño de los datos o de la fuente de los datos. En los datos encontramos el origen de toda la problemática y quizás el deber moral de que todos los participantes en el desarrollo e implementación de un sistema de ML verifiquen hacia atrás y hacia adelante el riesgo que genera su sistema.

Por ejemplo, Ifeoma Ajunwa¹⁹⁷, señala que, en el ámbito laboral, muchas empresas al ingresar piden los datos del trabajador, luego los utilizan y posteriormente impiden a los trabajadores conocer qué se ha hecho con sus datos. No saben cómo los evalúan para su productividad. En el ámbito laboral, se han detectado disparidades en la contratación y en la forma de interpretar datos para la toma de decisiones. Margaret Hu¹⁹⁸, explica el caso *Cambridge Analytica-Facebook* (que involucró a millones de usuarios que fueron liberados y explotados sin la debida autorización) analizando una variedad de soluciones legales y políticas que se han propuesto para mejorar la privacidad de la información (una reforma a la Comisión Federal de Comercio e imponer requisitos procesales para respetar el debido proceso a actores privados). Paul Prinsloo¹⁹⁹ examina el uso de cajas negras en la

¹⁹⁷ Ifeoma Ajunwa, “The ‘Black Box’ at Work”, en *Big Data & Society*, julio 2020, <https://doi.org/10.1177/2053951720938093>.

¹⁹⁸ Hu, Margaret. “Cambridge Analytica’s Black Box” en *Big Data & Society*, julio 2020. <https://doi.org/10.1177/2053951720938091>. [13 de 20 de julio de 2022] La autora comenta que el escándalo Cambridge Analytica-Facebook generó una preocupación generalizada sobre los métodos implementados para apuntar a los votantes a través de algoritmos de perfiles basados en datos de usuarios de Facebook. El escándalo finalmente condujo a una multa de \$ 5 mil millones impuesta a Facebook por la Comisión Federal de Comercio (FTC) en julio de 2019. Sin embargo, la acción de la FTC ha sido criticada por no abordar adecuadamente la privacidad y otros daños que emanan de la liberación de Facebook de aproximadamente 87 millones de datos de usuarios de Facebook, que fueron explotados sin la autorización. El ensayo resume la respuesta de la FTC y concluye que en la necesidad de explorar “protecciones del tipo de debido proceso dentro de las acciones de aplicación de las agencias reguladoras como la FTC”.

¹⁹⁹ Paul, Prinsloo, “Of ‘Black Boxes’ and Algorithmic Decision-Making in (Higher) Education – A Commentary” *Big Data & Society*, enero, 2020. <https://doi.org/10.1177/2053951720933994>. [20 de julio de 2022]

toma de decisiones algorítmicas en universidades para la planificación operativa y para las evaluaciones.

A lo largo de la década de 2010, los académicos exploraron la política y la sociología de los datos, su regulación (a través de leyes de privacidad y secreto comercial, entre muchas otras reglas) y su papel en la información y orientación de los responsables de la formulación de políticas. COVID-19 ha dramatizado la importancia de este trabajo. Los datos sanitarios de calidad han sido fundamentales para el éxito de los esfuerzos encaminados a “aplanar la curva”. En las organizaciones políticas menos afortunadas, los datos de salud se politizan, se ignoran o se manipulan.

A partir de enero de 2020, con la pandemia del COVID-19, los gobiernos han estado recopilando datos de ubicación de millones de usuarios y teléfonos inteligentes para encontrar patrones sobre cómo se propaga el virus. Los análisis, a menudo tienen algún ingrediente de opacidad, lo que limita la responsabilidad y trae consecuencias de discriminación y hacia los derechos de determinados grupos. Además de las preocupaciones obvias de privacidad, ya existe evidencia de que la recopilación y publicación de datos de ubicación durante la pandemia de COVID-19 está teniendo un impacto discriminatorio en los grupos minoritarios, replicando las desigualdades estructurales en lugar de disminuirlas.²⁰⁰

Desde los datos, su obtención o entrega, en la limpieza de los datos y la selección o creación de variables y en el modelado e implementación. En todo puede existir sesgos. Hay sesgos inherentes a los datos y sesgos que generan las personas que están construyendo el modelo de datos.²⁰¹ Se debe hacer la mitigación a través de diferentes técnicas ya sea *pre-processing*, *in-processing* o *post-processing*. Los expertos recomiendan también guardar los metadatos para

²⁰⁰ Thoreson, Ryan. (2020). *Covid-19 Backlash Targets LGBT People in South Korea Government Should Act to Prevent Discrimination*. Human Rights Watch. Cfr. <https://www.hrw.org/news/2020/05/13/covid-19-backlash-targets-lgbt-people-south-korea> [20 de julio de 2022]

²⁰¹ Saleiro, P., Kuester, B., Stevens, A., Anisfeld, A., Hinkson, L., J., & Ghani, R, “Aequitas: A Bias and Fairness Audit Toolkit”, 2018, *ArXiv*, [abs/1811.05577](https://arxiv.org/abs/1811.05577). *Aequitas – Center for Data Science and Public Policy* (datasciencepublicpolicy.org) [20 de julio de 2022]

una revisión futura para ubicar la responsabilidad algorítmica. Una herramienta recomendada, de código abierto, que ayuda a examinar, informar y mitigar la discriminación y el sesgo en los modelos de ML a lo largo del ciclo de vida es la aplicación IA *Fairness 360*.²⁰²

A veces, el mejor modelo no es el más eficiente y eficaz, sino el que toma en cuenta la equidad (disminuir el sesgo y sumar el ingrediente de la equidad). Es decir, la eficiencia y la eficacia deben ir de la mano de la equidad como hipótesis de selección del mejor modelo. Como, por ejemplo, algoritmos en donde el género, la etnia o la edad, influyen o algunos modelos que utilizan datos personales (por la privacidad). Siempre habrá tensiones y no hay una fórmula única, se deben hacer la manipulación de los modelos algorítmicos para ver cuál funciona mejor.

*Aequitas*²⁰³ es una *open framework* para auditoría de sesgos de código abierto para desarrolladores de aprendizaje automático, analistas y responsables de políticas para auditar modelos de aprendizaje automático por discriminación y sesgo; busca tomar decisiones informadas y equitativas en torno al desarrollo e implementación de herramientas predictivas de evaluación de riesgos. Existen métricas básicas de medición como *Equal Parity*, *Equal Opportunity* o *False Negative Parity*, *False Omission Rate*, basadas en: *True Negative* (verdadero negativo); *False Negative* (falso negativo), donde la primera variable es el valor real y la segunda la predicción. Fraude y salió fraude; no fraude y salió fraude; no fraude y no salió fraude; fraude y no salió fraude. Se puede emitir un reporte y presentarlo a la empresa o gobierno. Si el tomador de decisión tiene la información y no hace algo, es responsabilidad de la empresa o del gobierno.

Los impactos de los sesgos algorítmicos son un riesgo para la empresa, especialmente aquellos que no son explicables y, como hemos dicho, es imposible que un modelo de ML no produzca sesgos ya que, de manera natural, sea por el ser humano o por la propia máquina, al categorizar la información inevitablemente

²⁰² <https://aif360.mybluemix.net/> [20 de julio de 2022]

²⁰³ Saleiro, P., Kuester, B., Stevens, A., Anisfeld, A., Hinkson, L., London, J., & Ghani, R. (2018). *Aequitas: A Bias and Fairness Audit Toolkit*. *ArXiv*, *abs/1811.05577*. Consúltase también el sitio del framework: <http://aequitas.dssg.io/>

algo o alguien queda fuera (o dentro) según los resultados que se busquen obtener. Existe una propuesta para mitigar estos sesgos basada en la gestión de los sesgos relacionados con la intención que tiene la empresa de implementar un modelo de ML; los sesgos que surgen debido a la distribución de muestras utilizadas para la capacitación de ese modelo de ML; y, los que están presentes en muestras de entrada individuales. Aclarando que no existe una solución sencilla ni completa para el problema de sesgos.²⁰⁴

Muchos conjuntos de datos contienen sesgos que inevitablemente afectan a los sistemas de inteligencia artificial que usan estos mismos datos para la toma de decisiones. Incluso una limpieza cuidadosa de los datos protegidos y de *proxy* no eliminará por completo el sesgo histórico contenido en los datos. Sin embargo, Drew Roselli, Jeanna Matthews, and Nisha Talagala recomiendan implementar una combinación de procesos para mitigar el impacto del sesgo algorítmico en las predicciones actuales y reducir el impacto en las predicciones futuras. La combinación de procesos que sugieren es la siguiente:

- Monitoreo de datos para detectar anomalías y determinar si el conjunto de entrenamiento es apropiado.
- Análisis cuantitativo para justificar los datos sustitutos y tener en cuenta los ciclos de retroalimentación.
- Un proceso de revisión que garantiza la corrección de los datos introducidos.
- Un proceso de evaluación que verifica que las predicciones se basan en valores de características razonables.
- Mecanismos para explorar nuevos resultados e incorporar estos resultados en predicciones futuras para superar el sesgo histórico.²⁰⁵

Una solución está en las auditorías algorítmicas o en lo que se conoce como rendición de cuentas algorítmicas. La Tabla “A” contiene algunos problemas descritos y ofrece ciertas acciones de mitigación.

²⁰⁴ Drew Roselli, Jeanna Matthews y Nisha Talagala, “Managing Bias in AI” en *Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, 2019, pp. 539-544. DOI: <https://doi.org/10.1145/3308560.3317590>

²⁰⁵ *Idem*.

Problemas que causan sesgos y acciones de mitigación recomendadas

Tipo de problema	Asunto	Mitigaciones
Representando el problema	Objetivos proxy	Fundamentar la hipótesis con evidencia cuantitativa externa. Evaluar el impacto de las predicciones frente a métricas externas.
	Selección de características	Utilice conjuntos de herramientas específicos para ayudar a identificar sesgos ocultos. Validar las razones de la predicción como parte de la evaluación del modelo. Evalúe los casos no aprendidos en busca de pistas sobre los atributos que faltan.
	Datos sustitutos	Fundamentar los datos sustitutos señalando las limitaciones conocidas. Habilite un proceso de revisión de datos de entrada para ayudar a detectar limitaciones no obvias.
Conjuntos de datos	Casos invisibles	Detecta y revisa patrones de características de entrada inesperados.
	Conjuntos de datos no coincidentes	Evite los datos de entrenamiento demasiado seleccionados. Detecte cuándo los datos de entrada difieren de las distribuciones anticipadas durante el entrenamiento.
	Datos manipulados	Considere cómo se pueden manipular los datos de capacitación y los datos utilizados como entrada y habilitar salvaguardas o utilizar fuentes alternativas.
	Casos no aprendidos	Revise los casos que no se aprendieron durante la capacitación como parte de la evaluación del modelo.
	Funciones no generalizables	Durante la evaluación del modelo, determine qué características se utilizaron para hacer predicciones y compararlas con las utilizadas por un observador externo.
	Correlaciones irrelevantes	
	Problemas con los datos históricos	Se utilizaron conjuntos de herramientas específicos que ayudan a eliminar el sesgo histórico. Complemente los conjuntos de entrenamiento con muestras infrarrepresentadas. Utilice experimentos de aleatorización y A / B para explorar nuevos resultados.
Muestras individuales	Datos inexactos	Detecta datos incompletos. Habilite un proceso de revisión de datos de entrada para ayudar a detectar y corregir la entrada incorrecta.
	Datos obsoletos	Actualice los conjuntos de datos almacenados en caché con regularidad.

Fuente: Roselli, Drew, Matthews, Jeanna y Talagala, Nisha, "Managing Bias in AI" en *Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, 2019, pp. 539-544. DOI: <https://doi.org/10.1145/3308560.3317590> [13 de julio de 2022]

Ahora pasaremos al estudio de dos conceptos necesarios para entender el funcionamiento de los llamados algoritmos de caja negra y lo complicado que es la creación de una legislación que equilibre el desarrollo de este tipo de tecnologías y el crecimiento de la empresa sin afectar Derechos Humanos. La opacidad, nos lleva al estudio de dos conceptos nuevos para el Derecho, la interpretabilidad y explicabilidad algorítmicas como ingredientes para ubicar la responsabilidad algorítmica a través de acciones de rendición de cuentas algorítmicas (como lo son las auditorías algorítmicas):

B) De la opacidad a la interpretabilidad y explicabilidad algorítmicas

Respecto a la opacidad. En 2015, Frank Pasquale, profesor en la Facultad de Derecho de Brooklyn, publica el libro *The Black Box Society: The Secret Algorithms That Control Money and Information*. Su trabajo es considerado toral para entender el uso de algoritmos y las consecuencias que producen en la sociedad al momento de utilizar datos de las personas, sobre todo, vistos como instrumentos de poder financiero y explotación en la economía digital. Muchas empresas -dice el autor- están controlando nuestras vidas a través de la recopilación de datos y nuestro conocimiento sobre cómo funciona esto es parcial e incompleto. Es más, se puede considerar -en la mayoría de los casos- inaccesible ya que las empresas privadas custodian extremadamente su información y la hacen de imposible acceso para el propietario de ésta, para los académicos o para la sociedad en general.²⁰⁶

Frank Pasquale comenta que los burós de crédito “fueron pioneros en las técnicas de caja negra, emitiendo juicios críticos sobre las personas y ocultando sus métodos de recopilación y análisis de datos. En la década de 1960, las insinuaciones se filtraron en los informes presentados por “investigadores” no capacitados...”²⁰⁷.

²⁰⁶ Pasquale, Frank, *The Black Box Society: The Secret Algorithms. That Control Money and Information*, Cambridge, Harvard University Press, Massachusetts, 2015. Consúltese en: <https://doi.org/10.4159/harvard.9780674736061> [15 agosto de 2021]

²⁰⁷ Ibidem. p. 22.

¿Qué son los algoritmos de caja negra? Se dice que es de caja negra el algoritmo del que no es posible dar una interpretación de lenguaje natural y referida a los valores de las predicciones y del porqué el algoritmo toma determinada decisión. Pero primero, debemos comprender completamente el problema. Pasquale nos comenta que el término caja negra es una metáfora útil... dado su propio significado dual. “Puede referirse a un dispositivo de grabación, como los sistemas de monitoreo de datos en aviones, trenes y automóviles. O puede significar un sistema cuyo funcionamiento es misterioso; podemos observar sus entradas y salidas, pero no podemos decir cómo una se convierte en la otra.”²⁰⁸ El ejemplo más conocido de un algoritmo de caja negra lo tenemos en el buscador de Google²⁰⁹:

Google es quizás el caso más instructivo de cómo se desarrolló la cultura de la caja negra y por qué es importante. Antes de Google, la navegación web para los consumidores a menudo significaba portales abarrotados, anuncios incómodos y abundancia de spam. Google se hizo cargo del campo al ofrecer resultados claros, limpios y relevantes en fracciones de segundo. Incluso los escépticos de Silicon Valley le dan crédito a Google por poner orden en el caos. Para un investigador experto, Google es un regalo del cielo, una dinámica biblioteca alejandrina de contenido digital. Pero el éxito comercial le ha dado a la empresa un poder casi inconcebible, sobre todo sobre lo que encontramos en línea. Google no revela los detalles de sus métodos de clasificación. Ha explicado sus líneas generales y el proceso parece tranquilizadamente sencillo. Califica los sitios por relevancia e importancia. Cuantas más páginas web se vinculen a una página determinada, más autoritario lo considerará Google. (Para aquellos que necesitan conectarse a una página, pero no quieren promocionarla, Google promete no contar los enlaces que incluyan una etiqueta “rel:nofollow” tag). La votación está ponderada; Las páginas web que están vinculadas por muchas otras páginas tienen más autoridad que las no conectadas. Este es el núcleo del método patentado “PageRank” detrás del éxito de Google. El híbrido de PageRank de igualitarismo (cualquiera puede vincular) y elitismo (algunos vínculos cuentan más que otros) reflejan e inspiran modos poderosos de ordenar contenido web... Y así como el tiempo de guerra da a los gobiernos razones (y excusas) para ocultar sus planes al público, Google ha utilizado la batalla interminable contra el spam y la manipulación para justificar su negativa a dar cuenta de las controvertidas decisiones de clasificación.²¹⁰

²⁰⁸ Ibidem., p. 3.

²⁰⁹ Para conocer la historia del buscador de Google puede consultarse el artículo original de sus creadores Sergey Brin y Lawrence Page en: <http://infolab.stanford.edu/~backrub/google.html> [20 de julio de 2022]

²¹⁰ Pasquale, Frank, *op. cit.*, p. 64-65.

En todo ello, aparecen dos ingredientes, los datos y el uso de algoritmos para la toma de decisiones siendo la fórmula de caja negra. Las nuevas tecnologías como el *Big Data*, el cómputo en la nube y el internet de las cosas están generando para las empresas un volumen mayor de datos sobre sus clientes o usuarios, pero a la vez, las empresas luchan contra las regulaciones que permiten a los mismos usuarios ejercer cierto control sobre los expedientes digitales resultantes o sobre temas relacionados con la afectación de algún derecho humano (salud, trabajo, educación, etcétera).

Hubo en ese período un decremento de la privacidad de las personas y un incremento en los niveles de opacidad en el manejo de los datos que utilizan los gobiernos y las empresas, a pesar de que las normas que regulan el tema de la privacidad en el manejo de datos personales han incrementado su número, perfeccionamiento, rigurosidad y divulgación. Es muy probable que nuestros datos ya estén siendo utilizados cotidianamente en puntajes, clasificaciones, mediciones de riesgos o listas con alguna utilidad. Los algoritmos que generan esto lo hacen sin una supervisión externa o algún acto de fiscalización distinto al que pudiera hacerse al interior de la empresa; solo surge el control cuando existe un conflicto entre el usuario y la empresa, perdiendo ésta el control. Aquí es cuando descubrimos el destino de los datos y su tratamiento está alejado de sus dueños, quienes enfrentan un camino tortuoso para su defensa o protección.

Si bien es cierto que en ocasiones está justificada la secrecía de los algoritmos y la forma en cómo llegan a esos resultados, también lo es que en otras no existe una razón de peso más que la negativa de abrir el algoritmo y su mecanismo para llegar a determinada solución por una decisión de una persona o grupo de personas que se empeñan en ser transparentes (aun sabiendo que la información no les pertenece y deben seguir el procedimiento de protección que marca la ley).

Incluso con ese compromiso, no podemos olvidar que el acceso a los datos es solo el primer paso hacia la equidad y muchas de nuestras actividades diarias se procesan como “señales” en el mundo digital. No debemos olvidar también que las

decisiones críticas no se toman sobre la base de los datos *per se*, sino sobre la base de los datos analizados algorítmicamente: es decir, en cálculos codificados en un software para hacer predicciones.

La predicción a través de patrones es una capacidad de las máquinas que supera al ser humano, pero esa predicción genera sesgos que pueden discriminar a las personas. Los sesgos también generan problemas éticos que tienen que ver con la capacidad para identificar patrones y seguir comportamientos más allá de lo que antes era posible. En cuestiones de uso de vehículos sin conductor, está el clásico problema del carro que debe decidir entre causar un gran accidente o un accidente menor por un imprevisto en el camino. Si tiene que elegir entre toparse con una señal y lastimar a los pasajeros en el vehículo, o toparse con peatones al costado de la carretera, pero potencialmente salvar a los pasajeros (*Moral Machine*).

Es importante reiterar que uno de los problemas vinculados con el uso de algoritmos de IA basada en Machine Learning – Deep Learning lo encontramos en la opacidad y en los patrones oscuros que generan cierto tipo de algoritmos. Jenna Burrell²¹¹, considera el tema de la opacidad como un problema para los mecanismos de clasificación como filtros de spam, detección de fraudes con tarjetas de crédito, motores de búsqueda, tendencias de noticias, segmentación del mercado y publicidad, calificación de seguros o préstamos y calificación crediticia. Todos estos mecanismos de clasificación con frecuencia se basan en algoritmos de aprendizaje automático para realizar este trabajo. La autora hace una distinción sobre los tipos de opacidad existentes: 1) opacidad como secreto corporativo o estatal intencional, 2) opacidad como analfabetismo técnico y 3) una opacidad que surge de las características de los algoritmos de aprendizaje automático y escala requerido para aplicarlos de manera útil.

El objetivo de su propuesta se centra en profundizar en los algoritmos de aprendizaje automático y la naturaleza de su opacidad, relacionando esto con los intereses sociológicos de la clasificación y la discriminación. Esto es parte de una reorientación continua sobre la desigualdad digital que frecuentemente se ha

²¹¹ Jenna, Burrell, "How the Machine 'Thinks': Understanding Opacity in *Machine Learning Algorithms*" en *Big Data & Society*, junio 2016. <https://doi.org/10.1177/2053951715622512>. [20 de julio de 2022]

centrado en la distribución de recursos y habilidades computacionales, pero no, hasta hace poco, a la cuestión de cómo las personas pueden estar sujetas a clasificación computacional, invasiones de privacidad u otra vigilancia de maneras desiguales en la población general y que podrían violar las protecciones regulatorias existentes.

Las críticas legales de la opacidad algorítmica a menudo se centran en la capacidad de secreto intencional y conducen a regulaciones para hacer cumplir la transparencia. Autores como Pasquale abogan por el uso de auditores que tengan acceso al código y puedan asegurar que las clasificaciones no sean discriminatorias. Otro enfoque es educar a un sector más amplio de la sociedad en la escritura de códigos y habilidades computacionales para disminuir el problema de una clase homogénea y élite de personas técnicas que toman decisiones consecuentes que no pueden ser evaluadas fácilmente por los no miembros.

Sin embargo, la opacidad de los algoritmos de aprendizaje automático es un desafío en un nivel más fundamental. Cuando una computadora aprende y, en consecuencia, construye su propia representación de una decisión de clasificación, lo hace sin tener en cuenta la comprensión humana. Las optimizaciones de máquina basadas en datos de entrenamiento no concuerdan naturalmente con las explicaciones semánticas humanas.

Los ejemplos de reconocimiento de escritura a mano y filtrado de spam ayudaron a ilustrar cómo el funcionamiento de los algoritmos de aprendizaje automático puede escapar a la comprensión e interpretación total por parte de los humanos, incluso para aquellos con capacitación especializada, incluso para los informáticos.

En última instancia, las asociaciones entre académicos (legales, científicos sociales, expertos en dominios, junto con científicos informáticos), pueden eliminar estas desafiantes preguntas de equidad en la clasificación a la luz de la barrera de la opacidad. Además, las poblaciones de usuarios y el público en general pueden dar su voz respecto las exclusiones y formas de discriminación experimentada (algorítmica o de otro tipo) de las que los expertos en dominios pueden carecer de información.

No todos los sistemas de IA basada en Machine Learning – Deep Learning son cajas negras. Al igual que con los procesos de análisis del comportamiento humano o los modelos de algoritmos tradicionales, algunos sistemas de ML son simples y fáciles de explicar, mientras que otros son mucho más complejos. La interpretabilidad y la explicabilidad son dos áreas de investigación novedosas que están produciendo nuevos métodos y herramientas que nos dan una idea de por qué un sistema de IA se comporta de cierta manera.

Las explicaciones de las decisiones humanas, “por ejemplo, pueden no reflejar con precisión sus factores de influencia o sesgos inconscientes. De hecho, incluso si cada decisión individual tomada por algunos sistemas de IA no se puede explicar completamente, es posible que podamos comprender cómo toman decisiones en general mejor de lo que entendemos cómo los humanos toman decisiones similares.”²¹².

La interpretabilidad parte del concepto interpretar, que es una acción que todos los seres humanos hacen de forma recurrente. Sucede así porque el mundo, su actuar y el proceso de comunicación está lleno de signos, hechos o datos que requieren de interpretación. La interpretación “consiste en un acto por el cual se dota o asigna un significado específico a ciertos hechos, fórmulas, signos, acontecimientos o palabras”²¹³.

²¹² Cfr. Google, *Exploring 6 AI Myths*. Google, 2020. <https://ai.google/static/documents/exploring-6-myths.pdf>

²¹³ Rolando, Tamayo y Salmorán, *Razonamiento y argumentación jurídica. El paradigma de la racionalidad y la ciencia del derecho*, UNAM, 2003, p. 134. El filósofo que sentó las bases para analizar cómo el ser humano construye sus ideas, razona e interpreta es Aristóteles en su *Tratado de lógica (órganon)*, que se puede consultar en los dos tomos de Editorial Gredos (Madrid, 1982). También se recomiendan las obras de Emilio Betti *Teoría generale della Interpretazione*, Milano, Dott. A. Giuffrè Editore, 1955 y Hans-Georg Gadamer, *Verdad y método*, Tomos I y II, Ediciones Sígueme, Octava Edición, Salamanca 1999. En el ámbito jurídico, podemos hablar también, de hermenéutica y de la hermenéutica jurídica. La primera, según Mauricio Beuchot es “...Un instrumento de análisis y, sobre todo, de síntesis (comprehensiva) de los signos, que aporta el aspecto de interpretación, la cual pertenece a la operación mental de la comprensión y la explicación.” Cfr. *Elementos de Semiótica*, Editorial Surge, México, p. 300. La segunda, la sintetiza Gadamer cuando dice que “...ejerce una evidente función creadora; es una disciplina subsidiaria de la dogmática del derecho y busca el uso correcto del derecho ya que hay una tensión entre la generalidad de la legislación y la singularidad del caso concreto.”. Cfr. Gadamer, Hans-Georg, *Verdad y método*, Tomo II, Ediciones Sígueme, Octava Edición, Salamanca, 1999, pp. 72, 96, 106 y 108. La hermenéutica jurídica tiene varios métodos de interpretación que se utilizan por los juzgadores. En particular, el principio de “interpretación conforme” que es muy utilizado para resoluciones en las que se ven afectados Derechos Humanos tiene su origen en la jurisprudencia de

Existen muchos conceptos relacionados con la interpretabilidad que se pueden optimizar: Por ejemplo, las nociones de equidad o imparcialidad implican que los grupos protegidos (explícitos o implícitos) no son discriminados de alguna manera. Privacidad significa que el método protege la información confidencial de los datos. Propiedades como la fiabilidad y solidez determinan si los algoritmos alcanzan ciertos niveles de rendimiento frente a los parámetros o la variación de entrada. La causalidad implica que el cambio previsto en la producción debido a una perturbación ocurrirá en el sistema real.²¹⁴

En el contexto de los sistemas de ML, el ingrediente de la interpretabilidad (*interpretability*), parte de la siguiente hipótesis: Si el sistema puede explicar su razonamiento, entonces los seres humanos pueden verificar si ese razonamiento es correcto, completo, adecuado a/con respecto otros criterios, normas, reglas, derechos.

Visto esto como un principio que se debe intentar cumplir al usar IA y, principalmente algoritmos de ML, la interpretabilidad es la capacidad para explicar o presentar en términos comprensibles algo a un ser humano. Interpretar en el ambiente de la IA significa “explicar o presentar en términos comprensibles”²¹⁵. En el contexto de los sistemas de IA basada en Machine Learning – Deep Learning , definimos “la interpretabilidad como la capacidad de explicar o presentar en términos comprensibles a un humano”²¹⁶. Doshi-Velez, F., & Kim, B., al explicar el tema desde el enfoque de la psicología retoman las ideas de Tania Lombrozo²¹⁷ al

la Corte Suprema de los Estados Unidos, particularmente en el voto formulado por el juez Samuel Chase en el caso *Hylton vs United States* (1796). La Primera Sala de la SCJN ha dicho que el principio de interpretación conforme de todas las normas del ordenamiento a la Constitución se ve reforzado por el principio pro-persona, el cual obliga a maximizar la interpretación conforme en aquellos escenarios en los cuales, dicha interpretación permita la efectividad de los derechos fundamentales de las personas frente al vacío legislativo que puede provocar una declaración de inconstitucionalidad de la norma. Consúltase en: <https://sjf.scjn.gob.mx/SJFSem/Paginas/Reportes/ReporteDE.aspx?idius=27130&Tipo=2> [20 de julio de 2022]

²¹⁴ *Idem*.

²¹⁵ Doshi-Velez, F. y Kim, B., *Towards A Rigorous Science of Interpretable Machine Learning*. arXiv: *Machine Learning*, 2019, p. 2.

²¹⁶ *Idem*.

²¹⁷ Tania Lombrozo, “The structure and function of explanations”. *Trends in cognitive sciences*, 10(10): 464-470, citado por Doshi-Velez, F., y Kim, B. *op. cit.*, p. 2.

afirma que “las explicaciones son ... la moneda en la que intercambiamos creencias” y hacen una advertencia respecto de determinadas preguntas que se deben tener en mente: “¿qué constituye una explicación?, ¿qué hace que algunas explicaciones sean mejores que otras?, ¿cómo se generan las explicaciones?” Los investigadores clasificaron las explicaciones como “deductivas-nomológicas”, siguiendo a Hempel y Oppenheim (es decir, como pruebas lógicas).²¹⁸

Pero ¿qué podemos entender por el término explicación, ¿qué es algo explicable? Finale Doshi-Vélez y Kim, utilizando un texto de la psicóloga Tania Lombrozo, nos dicen:

“las explicaciones son... la moneda en la que intercambiamos creencias” y señala qué es lo que constituye una explicación, qué hace que algunas explicaciones sean mejores que otras, cómo se generan las explicaciones y cuándo se buscan explicaciones... Los investigadores han clasificado las explicaciones de ser “deductivos-nomológicos” en la naturaleza [Hempel y Oppenheim, 1948] (es decir, como pruebas lógicas) a proporcionar algún sentido... [Bechtel y Abrahamsen, 2005, Chater y Oaksford, 2006, Glennan, 2002]. Keil [2006] ... proponemos formas basadas en datos para obtener definiciones operativas y evaluaciones de explicaciones y, por lo tanto, de interpretabilidad.²¹⁹

Los sistemas de IA deben cumplir siempre los principios de: 1) responsabilidad, 2) equidad, 3) gobernabilidad, 4) precaución, 5) rendición de cuentas, 6) imputabilidad, 7) previsibilidad, 8) trazabilidad, 9) seguridad de funcionamiento, 10) fiabilidad, 11) transparencia, 12) explicabilidad y 13) proporcionalidad.²²⁰

Adicionalmente, la interpretabilidad, como herramienta para explicar la IA, articula todos estos principios para proteger Derechos Humanos. Podríamos decir que es el principio de principios de la IA y va de la mano de la explicabilidad. Por

²¹⁸ Doshi-Velez, F., & Kim, B., p. 2.

²¹⁹ Lombrozo T., *op. cit.*, en Doshi-Velez, F., & Kim, B., *op. cit.*, p. 2. La traducción es propia.

²²⁰ El ser humano ha sido insistente en poner a los sistemas de IA principios para proteger a su persona, a otras o a la sociedad en conjunto. Las Tres Leyes de la Robótica, quizás sean las más famosas: “1. Un robot no debe dañar a un ser humano, por su inacción, dejar que un ser humano sufra daño. 2. Un robot debe obedecer las órdenes que le son dadas por un ser humano, excepto cuando estas órdenes se oponen a la Primera Ley. 3. Un robot debe proteger su propia existencia hasta donde esa protección no entre en conflicto con la Primera o la Segunda Ley.” Cfr. Isaac Asimov, *Yo robot*, segunda edición, México, 2016.

ejemplo, al momento que se busca que una máquina produzca algún código que le permita al ser humano realizar su interpretación van apareciendo preguntas relacionados con el principio de equidad o imparcialidad que buscan proteger a grupos vulnerables (no solo por ser discriminados); nos encontramos también con el principio de privacidad, que obliga revisar si el algoritmo protege la información confidencial en los datos de las personas; está el principio la fiabilidad de los algoritmos y su rendimiento frente a parámetros o variaciones de entrada; y, también se relaciona con el principio de causalidad, es decir, verificar y validar que el cambio previsto en la materialización de un siniestro en el sistema sea real.

A medida que los sistemas de aprendizaje automático se vuelven omnipresentes, ha surgido un gran interés en el aprendizaje automático interpretable: sistemas que brindan una explicación de sus resultados. Estas explicaciones se utilizan a menudo para evaluar cualitativamente otros criterios como la seguridad o la no discriminación.

Finale Doshi-Velez y Been Kim llegan a la conclusión de que no todos los sistemas de IA basada en Machine Learning – Deep Learning requieren interpretabilidad. Es decir, no todo requiere una explicación porque 1) en muchos casos no existen consecuencias significativas para resultados inaceptables o 2) el problema está suficientemente bien estudiado y validado en aplicaciones reales en las que la decisión del sistema goza de gran credibilidad.

Entonces: ¿Cuándo es necesaria la interpretabilidad? Doshi-Velez y Been Kim argumentan que en términos de ML se debe diferenciar entre los casos en los que las incógnitas dan lugar a variaciones cuantificadas (por ejemplo, tratar de aprender de un pequeño conjunto de datos o con sensores limitados) y la incompetencia que produce algún tipo de sesgo no cuantificado, por ejemplo, el efecto de incluir el conocimiento del dominio en un proceso de selección de modelos de ML. La necesidad de interpretabilidad por una incompetencia en la formalización del problema crea un obstáculo fundamental para la optimización y la evaluación del modelo de ML. Para ejemplificar esto, Doshi y Kim muestran algunos casos relacionados con:

- Comprensión científica: El objetivo del ser humano es adquirir conocimiento. Sin embargo, no tenemos una forma completa de decir qué es el conocimiento; por lo tanto, lo mejor que podemos hacer es pedir explicaciones que podamos convertir en conocimiento.
- Seguridad: Para tareas complejas, el sistema de extremo a extremo casi nunca es completamente comprobable; no se puede crear una lista completa de escenarios en los que el sistema puede producir un error. Enumerar todas las salidas posibles dados todas las entradas posibles es computacional o logísticamente inviable, y es posible que no podamos marcar todas las salidas indeseables.
- Ética: El ser humano puede querer protegerse contra ciertos tipos de discriminación, y su noción de justicia puede ser demasiado abstracta para ser completamente codificada en el sistema (por ejemplo, uno podría desear un clasificador “justo” para la aprobación del préstamo). Incluso si podemos codificar protecciones para clases protegidas específicas en el sistema, podría haber sesgos que no considerábamos a priori (por ejemplo, uno puede no construir incrustaciones de palabras sesgadas por género a propósito, pero fue un patrón en los datos que se hizo evidente sólo después del hecho).
- Objetivos no coincidentes: El algoritmo del agente puede estar optimizando un objetivo incompleto, es decir, una función de proxy para el objetivo final. Por ejemplo, un sistema clínico puede optimizarse para el control del colesterol, sin tener en cuenta la probabilidad de adherencia; un ingeniero automotriz puede estar interesado en los datos del motor no para hacer predicciones sobre fallas en el motor, sino para construir más ampliamente un coche mejor.
- Compensaciones multi-objetivos: Dos desideratas bien definidas en los sistemas de ML pueden competir entre sí, como la privacidad y la calidad de predicción [Hardt et al., 2016] o privacidad y no discriminación [Strahilevitz, 2008]. Incluso si cada objetivo se especifica plenamente, es posible que no se conozca plenamente la dinámica exacta de la compensación y que la decisión tenga que ser caso por caso.²²¹

Respecto de la Explicabilidad. La interpretabilidad nos lleva a la necesidad de poder hacer una explicación de lo ocurrido (al empresario, al usuario, al gobierno en sus funciones de supervisión), es decir: a la explicabilidad algorítmica. Entre los métodos que utiliza la IA basada en Machine Learning – Deep Learning para hacer una explicación y los tipos de explicaciones que se les exigen a sus creadores por

²²¹ Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. arXiv: *Machine Learning*, p. 2. La traducción es propia. En presencia de una incompetencia, las explicaciones son una de las maneras de garantizar que los efectos de las lagunas en la formalización de problemas sean visibles para nosotros. Según los autores: “Existen 3 propuestas de clasificación de las evaluaciones de la interpretabilidad: Evaluación basada en aplicaciones (humanos reales, tareas reales); Métricas basadas en humanos (humanos reales, tareas simplificadas); y, Evaluación basada funcionalidad (sin humanos, tareas de representación). La evaluación basada en la funcionalidad no requiere experimentos humanos; en su lugar, utiliza alguna definición formal de interpretabilidad como un “proxy” para la calidad de la explicación. Estos experimentos son atractivos porque incluso los experimentos generales con sujetos humanos requieren tiempo y costos tanto para realizar como para obtener las aprobaciones necesarias (por ejemplo, IRBs), que pueden estar más allá de los recursos de un investigador de aprendizaje automático. Las evaluaciones basadas en la funcionalidad son las más apropiadas una vez que tenemos una clase de modelos o regularizadores que ya han sido validados, por ejemplo, a través de experimentos basados en humanos. También pueden ser apropiados cuando un método aún no está maduro o cuando los experimentos con sujetos humanos no son éticos. El reto, por supuesto, es determinar qué representantes utilizar. Por ejemplo, los árboles de decisión se han considerado interpretables en muchas situaciones.”. La traducción es propia. Cfr. Doshi-Velez, F., & Kim, B. (2017) pp. 5 y 6.

el usuario o la sociedad, no existe una conexión como tal. Este es un primer obstáculo, visto esto desde las diferentes demandas o exigencias sociales o, de determinados grupos y gobiernos. Es decir, por un lado, están los llamados métodos de inteligencia artificial explicativa (*explanatory artificial intelligence, XAI*) y, por el otro, los tipos de explicaciones que exige la sociedad y que le son de utilidad. Lo sorprendente es que ambos se encuentran separados, sin una vinculación y solo hay algunos intentos de unión.

¿Qué es lo que trae esta situación de alejamiento y cómo se produce? Esto se presenta sobre todo al usar modelos de caja negra o modelos opacos, como por ejemplo aquellos basados en Redes Neuronales Profundas (RNP). Siguiendo el trabajo hecho por Leilani H. Gilpin, Cecilia Testart, Nathaniel Fruchter, Julius Adebayo²²² debemos desglosar el tema en varios elementos para una mejor comprensión:

- a) Una explicación de un modelo de IA basada en Machine Learning – Deep Learning debe tener dos ingredientes: a) ser interpretable y b) ser completa. La suma de estos elementos la hará técnicamente adecuada. Por interpretable, debe entenderse: que la explicación debe ser comprensible para los humanos. Eso no implica necesariamente que la explicación deba estar en forma legible por humanos, de hecho, las señales visuales son bien entendidas por los humanos. Y cuando decimos la explicación es completa, se refiere a que la explicación resultante debe ser fiel al modelo. Por ejemplo, si bien el uso de un modelo simplificado que es explicable (como un modelo lineal) para ajustar la entrada a la salida da como resultado una buena explicación, no es una representación completa de los conceptos, representaciones y decisiones internas del modelo.²²³
- b) Las explicaciones basadas en algoritmos de ML pueden clasificadas como internas y externas (para explicar las redes neuronales profundas). Son explicaciones internas, las que utilizan los propios creadores del algoritmo o

²²² Gilpin, L.H., Testart, C., Fruchter, N., y Adebayo, J., *Explanations to Society*, ArXiv, [abs/1901.06560](https://arxiv.org/abs/1901.06560). [1901.06560v1.pdf](https://arxiv.org/pdf/1901.06560v1.pdf) (arxiv.org) 2019. [20 de julio de 2022]

²²³ *Idem*.

los técnicos que las utilizan y es un tipo de explicaciones que existe desde que se comenzaron a crear los algoritmos de ML, es decir, que están dirigidas a los desarrolladores y expertos en IA, quienes saben de álgebra, estadística, probabilidad, etcétera. Están diseñados para expertos en este campo del conocimiento. Por el contrario, las explicaciones externas (interpretables, completas y respondan a preguntas como: ¿por qué esta entrada conduce a esta salida en particular? se deben construir casuísticamente para generar confianza no solo en sus desarrolladores técnicos, sino también en aquellas personas alejadas del ámbito técnico que utilizan la tecnología sin conocimientos especializados (el usuario, cliente, estudiante, médico, conductor, etcétera). Esta interpretación permite atender las explicaciones que exigen los gobiernos y la sociedad.²²⁴

Leilani H. Gilpin, Cecilia Testart, Nathaniel Fruchter, Julius Adebayo han encontrado que la política de los gobiernos y las asambleas legislativas en el mundo están buscando crear normas jurídicas para regular los sistemas de IA y proteger los derechos de las personas, de determinados grupos, las finanzas, el empleo o asuntos de interés público.²²⁵ Sin embargo, existe una desconexión entre los métodos de inteligencia artificial explicativa (XAI) -ya que son de tipo interno- y los tipos de explicaciones que son exigidas por la sociedad y los gobiernos, -ya que son externos- Las preguntas que los expertos hacen a los sistemas opacos proporcionan “explicaciones internas”, enfocado en su depuración, confiabilidad y validación.

El deber de hacer explicaciones externas lo encontramos en diversas normas jurídicas como el caso del Reglamento General de Protección de Datos (RGPD) de la Unión Europea, el cual crea obligaciones para los procesos de toma de decisiones automáticos y el derecho a una explicación (explicabilidad algorítmica). La consideración 71 expresa lo siguiente:

²²⁴ *Idem.*

²²⁵ *Idem.*

El interesado debe tener derecho a no ser objeto de una decisión, que puede incluir una medida, que evalúe aspectos personales relativos a él, y que se base únicamente en el tratamiento automatizado y produzca efectos jurídicos en él o le afecte significativamente de modo similar, como la denegación automática de una solicitud de crédito en línea o los servicios de contratación en red en los que no medie intervención humana alguna. Este tipo de tratamiento incluye la elaboración de perfiles consistente en cualquier forma de tratamiento de los datos personales que evalúe aspectos personales relativos a una persona física, en particular para analizar o predecir aspectos relacionados con el rendimiento en el trabajo, la situación económica, la salud, las preferencias o intereses personales, la fiabilidad o el comportamiento, la situación o los movimientos del interesado, en la medida en que produzca efectos jurídicos en él o le afecte significativamente de modo similar. Sin embargo, se deben permitir las decisiones basadas en tal tratamiento, incluida la elaboración de perfiles, si lo autoriza expresamente el Derecho de la Unión o de los Estados miembros aplicable al responsable del tratamiento, incluso con fines de control y prevención del fraude y la evasión fiscal, realizada de conformidad con las reglamentaciones, normas y recomendaciones de las instituciones de la Unión o de los órganos de supervisión nacionales y para garantizar la seguridad y la fiabilidad de un servicio prestado por el responsable del tratamiento, o necesario para la conclusión o ejecución de un contrato entre el interesado y un responsable del tratamiento, o en los casos en los que el interesado haya dado su consentimiento explícito. En cualquier caso, dicho tratamiento debe estar sujeto a las garantías apropiadas, entre las que se deben incluir la información específica al interesado y el derecho a obtener intervención humana, a expresar su punto de vista, a **recibir una explicación de la decisión tomada después de tal evaluación y a impugnar la decisión**. Tal medida no debe afectar a un menor.

A fin de garantizar un tratamiento leal y transparente respecto del interesado, teniendo en cuenta las circunstancias y contexto específicos en los que se tratan los datos personales, el responsable del tratamiento debe utilizar procedimientos matemáticos o estadísticos adecuados para la elaboración de perfiles, aplicar medidas técnicas y organizativas apropiadas para garantizar, en particular, que se corrijan los factores que introducen inexactitudes en los datos personales y se reduce al máximo el riesgo de error, asegurar los datos personales de forma que se tengan en cuenta los posibles riesgos para los intereses y derechos del interesado y se impidan, entre otras cosas, efectos discriminatorios en las personas físicas por motivos de raza u origen étnico, opiniones políticas, religión o creencias, afiliación sindical, condición genética o estado de salud u orientación sexual, o que den lugar a medidas que produzcan tal efecto. Las decisiones automatizadas y la elaboración de perfiles sobre la base de categorías particulares de datos personales únicamente deben permitirse en condiciones específicas.²²⁶

Esta consideración coloca una carga de responsabilidad sobre aquellos que procesan datos o desarrollan sistemas de IA para generar explicaciones externas. También lleva a la reflexión sobre las preocupaciones sociales y gubernamentales relacionadas con el uso de la IA basada en Machine Learning – Deep Learning. Por ejemplo: ¿Cómo puede saber que el banco tomó la decisión correcta en el otorgamiento de un préstamo/crédito? ¿Cómo saber si el médico está tomando la

²²⁶ Para mayores datos puede consultarse: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679&from=ES> [20 de julio de 2022]

mejor decisión al utilizar un sistema de ML para emitir su diagnóstico? ¿Qué garantiza que la información que da soporte a estas decisiones es interpretable? ¿El sistema podía tomar una decisión diferente sin ser explicable?

Leilani H. Gilpin, Cecilia Testart, Nathaniel Fruchter, Julius Adebayo afirman que los algoritmos de IA y los sistemas de IA más generales y complejos actualmente no pueden proporcionar respuestas a este tipo de preguntas ya que no están diseñados para explicar al público en general o a los responsables del gobierno cómo procesan los datos para llegar a una resolución. Aunque ya ha habido extrañamientos a las empresas para modificar esto, los sistemas explicativos que existen actualmente fueron pensados para el programador o experto, no para un usuario final o un responsable del gobierno (solo producen explicaciones internas). Cuando ocurre un accidente con algún sistema de IA basada en Machine Learning – Deep Learning los usuarios, responsables del gobierno, aseguradoras, peritos, jueces, fiscales, van a desear saber quién o qué es el responsable del accidente y por qué ocurrió y; se generará un vacío para la toma de decisiones.²²⁷

Leilani H. Gilpin, Cecilia Testart, Nathaniel Fruchter, Julius Adebayo hacen una propuesta de los tipos de preguntas que los algoritmos RNP explicativos pueden y no pueden responder, centrándose en las RNP, debido al cambio reciente en la investigación de IA de enfoques simbólicos al aprendizaje automático y al aprendizaje profundo, y porque estos son los sistemas que toman decisiones críticas para la seguridad en aplicaciones como la conducción autónoma y la detección de malware.²²⁸ La propuesta del equipo de expertos se puede apreciar en la siguiente tabla:

²²⁷ *Idem.*

²²⁸ *Idem.*

Fortalezas, beneficios y desafíos de los sistemas RNP XAI actuales

Método	Preguntas que SÍ puede responder	Preguntas que NO puede responder
Procesamiento	¿Por qué esta entrada en particular conduce a esta salida en particular?	¿Por qué estas entradas eran más importantes para la salida? ¿Cómo se podría cambiar la salida?
Representación	¿Qué información contiene el sistema?	¿Por qué una representación es relevante para las salidas? ¿Cómo se aprendió esta representación?
Producir una explicación	Dada una salida en particular o decisión, ¿cómo puede el sistema explicar su comportamiento?	¿Qué información contribuyó a esto? producción/decisión? ¿Cómo puede el sistema producir una salida/decisión diferente?

Fuente: Gilpin, L.H., Testart, C., Fruchter, N., y Adebayo, J., *Explanations to Society*, ArXiv, abs/1901.06560. [1901.06560v1.pdf \(arxiv.org\)](https://arxiv.org/pdf/1901.06560v1.pdf) 2019. [13 de julio de 2022]

La explicabilidad, comenta el Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial creado por la Comisión Europea, es un “principio ético crucial”²²⁹ para conseguir que los usuarios confíen en los sistemas de IA y para mantener dicha confianza. La finalidad de los sistemas de IA y que las decisiones deben poder explicarse —en la medida de lo posible— a las partes que se vean afectadas por ellas de manera directa o indirecta. Sin esta información, no es posible impugnar adecuadamente una decisión.”.²³⁰

Son los algoritmos basados en Redes Neuronales Profundas (RNP) los que requieren atención especial cuando se busca hacer su explicación. En estos casos, es necesario adoptar medidas distintas para alcanzar la explicabilidad (trazabilidad, auditabilidad y transparencia), siempre y cuando el sistema en su conjunto respete los Derechos Humanos. El grado de explicabilidad de las circunstancias y del contexto o la gravedad derivadas de un resultado equivocado o no esperado.

El Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial creado por la Comisión Europea hace una “Lista de Evaluación para una IA Fiable”:

²²⁹ Para el Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial Creado por la Comisión Europea existen cuatro principios éticos que siguen el orden de aparición de los derechos fundamentales que se basan en la Carta de la Unión Europea: 1. Respeto de la autonomía humana; 2. Prevención del daño; 3. Equidad; y 4. Explicabilidad.

²³⁰ Comisión de Bruselas, Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial Creado por la Comisión Europea, *Directrices Éticas para una IA fiable*, Bruselas, 2019, p. 22.

1. ¿Ha evaluado en qué medida son comprensibles las decisiones y, por tanto, el resultado producido por el sistema de IA?
2. ¿Se ha asegurado de que se pueda elaborar una explicación comprensible para todos los usuarios que puedan desearla sobre las razones por las que un sistema adoptó una decisión determinada que diera lugar a un resultado específico?
3. ¿Ha evaluado en qué medida la decisión del sistema influye en los procesos de adopción de decisiones de la organización?
4. ¿Ha evaluado por qué se desplegó ese sistema en particular en esa área concreta?
5. ¿Ha evaluado el modelo de negocio del sistema (por ejemplo, de qué modo crea valor para la organización)?
6. ¿Ha diseñado el sistema de IA teniendo en mente desde el principio la interpretabilidad?
7. ¿Ha investigado y tratado de utilizar el modelo más sencillo e interpretable posible para la aplicación en cuestión?
8. ¿Ha evaluado si puede analizar sus datos relativos a la formación y los ensayos realizados? ¿Puede modificar y actualizar estos datos a lo largo del tiempo?
9. ¿Ha evaluado si, tras la formación y el desarrollo del modelo, tiene alguna posibilidad de examinar su interpretabilidad o si dispone de acceso al flujo de trabajo interno del modelo? ²³¹

Como seres humanos, nos genera confianza hacer preguntas y saber sus respuestas, por ello es indispensable poder juzgar el comportamiento de los algoritmos RNP haciendo preguntas siguiendo la construcción del modelo y las exigencias del gobierno o la sociedad (caso por caso).

Para confiar verdaderamente en los sistemas de IA, las personas no solo necesitarán sentirse seguras de que comprenden con certeza cómo se toman las decisiones, sino también de que tienen recursos legales en caso de incumplimiento. Si una persona no está de acuerdo con la salida de un sistema, debe tener la posibilidad de cambiar el sistema, tal como lo prevé el RGDP.

Diseñar una IA explicable es importante, pero coincidimos con la conclusión que hacen Leilani H. Gilpin, Cecilia Testart, Nathaniel Fruchter, Julius Adebayo cuando dicen: "...solo cuando los sistemas opacos sean auditables, explicables, respondan preguntas e interpreten los comentarios, estaremos lo suficientemente seguros para confiar en su toma de decisiones".²³² Por el momento, la formación de equipos transdisciplinarios que participan con sistemas de revisión de la

²³¹ Ibidem., pp. 37-38.

²³² Gilpin, L.H., Testart, C., Fruchter, N., y Adebayo, J., *Explanations to Society*, ArXiv, [abs/1901.06560](https://arxiv.org/abs/1901.06560). [1901.06560v1.pdf](https://arxiv.org/pdf/1901.06560v1.pdf) (arxiv.org) 2019.

trazabilidad, auditabilidad y transparencia algorítmica, son una salida que se debe de considerar en las empresas y gobiernos para atender este tema.

Generalmente, -dicen los expertos citados- explicar el comportamiento del modelo no es suficiente para generar confianza. A veces el modelo puede comportarse de forma errónea debido a alguna inconsistencia en los datos de entrenamiento y / o bases de datos. Esto no significa necesariamente que los datos de entrenamiento sean malos, pero sí que hay algo desajustado entre los datos que dieron el resultado y los datos de entrenamiento utilizados. Un ejemplo claro es el caso de un algoritmo de reclutamiento de Amazon que generaba resultados extremadamente sesgados. El algoritmo se había entrenado con datos de solicitantes (donde los hombres eran dominantes) y, por tanto, se estaba entrenando a sí mismo para elegir candidatos masculinos.²³³

Otro caso es el uso de estos algoritmos para diagnosticar el cáncer de mama a partir de una imagen, donde el conjunto de entrenamiento se seleccionó cuidadosamente para estar bastante cerca de una división 50-50 de escaneos de cáncer de mama y no cáncer de mama. Incluso si el clasificador es muy preciso, sin tener acceso a explicaciones completas para comprender cómo se toman las decisiones en el modelo, no es seguro que esté tomando decisiones por las razones correctas.²³⁴

Joy Lu, Dokyun Lee, Tae Wan Kim y David Danks se preguntan cuál es una buena explicación para el resultado del aprendizaje automático y hacen un planteamiento interesante en tres fases: Primero, establecen la definición de una buena explicación desde las perspectivas de la Filosofía y la psicología. En particular, las teorías pragmáticas de explicación en la Filosofía de la ciencia y las teorías sobre las funciones cognitivas de la explicación en psicología sostienen que una explicación debería permitir al receptor de la explicación avanzar sus metas u objetivos. Estos objetivos pueden variar según el contexto, pero generalmente se centran en aumentar la comprensión, la confianza y la adopción entre los usuarios.

²³³ Ibidem., p. 4.

²³⁴ *Idem*

En segundo lugar, desarrollan un marco teórico para las características potencialmente relevantes de una explicación, así como los factores contextuales y / o situacionales, que pueden afectar la percepción del receptor de la explicación y, en última instancia, los objetivos. En tercer lugar, para demostrar cómo este marco puede adaptarse a un contexto del mundo real, realizan un experimento de laboratorio utilizando datos y decisiones financieras del mundo real para medir la efectividad de los diferentes modos de explicación y el impacto de las diferencias situacionales en el usuario -objetivos enfocados como la comprensión y la confianza.²³⁵

Las teorías de la explicación en Filosofía de la ciencia se dividen en dos grupos dependiendo de si los factores pragmáticos son necesariamente parte de lo que hace que algo sea una buena explicación. Las teorías tradicionales sostienen que algo es una buena explicación solo cuando proporciona la respuesta correcta a una pregunta de por qué (Strevens ofrece una visión general muy útil). Por ejemplo, una explicación de por qué alguien compró la pluma A en lugar de la pluma B podría proporcionar la secuencia causal que comienza con la decisión de comprar una pluma y culmina en la decisión final de compra.²³⁶

Las diferentes teorías explicativas resultan de diversas suposiciones o restricciones sobre qué aspectos del mundo pueden o deben incluirse en una explicación (por ejemplo, todos los factores causales, solo aquellas causas que pueden manipularse, hechos matemáticos, etc.). Todas estas teorías explicativas sostienen que la calidad de una explicación está determinada en última instancia por su correspondencia con el aspecto del mundo que se explica. Como resultado, la audiencia prevista de una explicación particular no juega ningún papel en la evaluación de su calidad; Si la audiencia destinataria no puede entender la explicación (por ejemplo, una explicación técnicamente precisa que involucra la

²³⁵ Joy Lu, Dokyun Lee, Tae Wan Kim, and David Danks, "Good Explanation for Algorithmic Transparency. Proceedings of the AAAI/ACM" *Conference on AI, Ethics, and Society*, Association for Computing Machinery, New York, número 93, en DOI: <https://doi.org/10.1145/3375627.3375821> [20 de julio de 2022]

²³⁶ *Idem.*

mecánica cuántica que se le da a un niño de cinco años), entonces el fallo es con la audiencia, no con la explicación.²³⁷

En contraste, las teorías pragmáticas de explicación, argumentan que una explicación debe proporcionar una respuesta a una pregunta de por qué y también le permite al receptor avanzar sus metas o intereses. Si alguien no puede entender una explicación propuesta o si la explicación no proporciona la información que necesita para sus objetivos o propósitos actuales, la culpa no recae en el destinatario; más bien, la explicación en sí misma se considera pobre (para esa audiencia). Por supuesto, diferentes audiencias pueden tener diferentes bases de conocimiento u objetivos, por lo que la misma explicación podría considerarse buena para una audiencia, pero no para otra. Es decir, las teorías pragmáticas de explicación implican pluralismo explicativo, o la idea de que puede haber múltiples buenas explicaciones para el mismo evento o por qué la pregunta, y esas explicaciones no necesitan ser consistentes con otro. La última condición es importante, ya que incluso los puntos de vista tradicionales reconocen que puede haber múltiples explicaciones de un evento, ya que podemos dar caracterizaciones precisas del mundo a diferentes niveles (es decir, macroeconomía vs. microeconomía vs. psicología vs. neurobiología) y todo lo de esas explicaciones pueden ser consistentes entre sí. El pluralismo explicativo es la idea más fuerte de que una buena explicación del porqué una persona podría no ser una buena explicación para otra persona, y es posible que no podamos unificar las dos explicaciones diferentes.²³⁸ Esto pasa seguido en medicina al generar diagnósticos/tratamientos o interpretaciones de estudios.

A continuación, expondremos el panorama de la rendición de cuentas algorítmicas que se ha venido construyendo durante los últimos años con legislación que ya está en proceso de aprobación o que ya está vigente y su relación con la transparencia y privacidad, y, por otro lado, la auditabilidad algorítmica como

²³⁷ *Idem.*

²³⁸ *Idem.*

instrumento adicional para la medición de riesgos y con el fin de evitar afectaciones a los Derechos Humanos de las personas.²³⁹

2) Rendición de cuentas algorítmicas

Los algoritmos y sus resultados están siendo analizados y regulados por organismos nacionales y supranacionales, por académicos y organizaciones de la sociedad civil; por eticistas, filósofos, ingenieros y científicos de datos que están siendo contratados por las grandes empresas de tecnología para guiar el camino de sus innovaciones. Las grandes empresas de tecnología están echando la mano de la Ética y a la ciencia de datos ya sea por gusto o por necesidad. Es un hecho inminente la regulación de la IA basada en Machine Learning – Deep Learning. Si bien al principio había cierta resistencia y se pensaba que eran suficientes los programas de integridad corporativa, a partir de 2018, diversos hechos cambiaron esta perspectiva y hoy las propias empresas exigen la intervención del gobierno y de sus facultades de control-legislación y control-supervisión. IBM, Google, Microsoft, Amazon, están en el Congreso americano cabildeando para que su experiencia se tome en cuenta y esa legislación se construya con base en las necesidades del mercado. Otros países, como Francia, han decidido irse directamente a la construcción de estructuras de gobierno para la supervisión de los algoritmos. Tal es el caso del Ministerio de Economía francés que ha creado un área especializada para ello y regular el uso de las grandes plataformas (Google, Amazon, Facebook y Apple).

Para tener éxito con los sistemas de IA basada en Machine Learning – Deep Learning se requiere entender y trabajar con las reglas de la privacidad y las reglas de la transparencia de la información. Se requiere entender el uso (o mejor dicho, mal uso) de los datos de las personas, la forma en cómo se deben de tratar, de cuidarse desde su almacenamiento y resguardo hasta su destrucción. En la siguiente sección explicaremos estos temas.

²³⁹ Consúltense también: Aizenberg, Evgeni y Jeroen van den Hoven. “Designing for Human Rights in AI”, *Big Data & Society*, julio 2020, doi:[10.1177/2053951720949566](https://doi.org/10.1177/2053951720949566). [20 de julio de 2022]

A) Transparencia y privacidad

Respecto a la transparencia. En 2021 se celebra el 255° aniversario de la “libertad de información” o “acceso a la información” en Suecia y Finlandia. Aquí encontramos la primera ley sobre libertad de información o acceso a la información aprobada por el parlamento sueco en 1766: *Freedom of Information Act (principle of publicity)*. Anders Chydenius (1729-1803), político finlandés, es considerado como el artífice de esta ley. La ley detiene la censura política y da al ciudadano la obtención del acceso público a todo documento gubernamental. El legado de Chydenius ha sido incluido por el Derecho Internacional en la Declaración Universal de Derechos Humanos, en el Pacto Internacional de Derechos Civiles y Políticos, en la Convención Americana de los Derechos Humanos y en la legislación de muchos estados (toda persona tiene derecho a acceder a la información pública).²⁴⁰

El artículo sexto proclama el principio de transparencia como parte de la libertad de prensa al declarar que “...esta libertad incluirá todos los intercambios de correspondencia, relaciones fácticas, documentos, protocolos, resoluciones judiciales y reconocimientos, ya se hayan redactado en el pasado o sean iniciados, mantenidos, presentados, dirigidos y concluidos en lo porvenir, antes, durante y después de los procedimientos ante los tribunales inferiores, de apelación y superiores y los departamentos gubernamentales, nuestros altos servidores y los consistorios y otros organismos públicos, y sin distinción entre la naturaleza de los casos, ya sean estos civiles, eclesiásticos o que de algún otro modo conciernan a controversias religiosas”²⁴¹

²⁴⁰ Anders Chydenius Foundation, *The World's First Freedom of Information Act*. Anders Chydenius Foundation Publications 2, s/l, 2016. https://www.chydenius.net/tiedostot/worlds_first_foia.pdf [20 de julio de 2022]

²⁴¹ El artículo sexto dice: “§6. This freedom of the press will further include all exchanges of correspondence, *species facti*, documents, protocols, judgments and awards, whether they were produced in the past or will be initiated, maintained, presented, conducted and issued hereafter, before, during and after proceedings before lower courts, appeal and superior courts and government departments, our senior administrators and consistories or other public bodies, and without distinction between the nature of the cases, whether these are civil, criminal or ecclesiastical or otherwise in some degree concern religious controversies; as well as older and more recent appeals and expositions, declarations and counter-declarations that have been or will be submitted to the Chambers of Our Supreme Court as well as the official correspondence and memorials that have

La ley contenía un límite a esa transparencia y tenía que ver con “las materias afectadas por el secreto ministerial, así como los informes y declaraciones realizadas sobre peticiones y apelaciones elevadas a los Estados del Reino” (artículo 8). Todos los “súbditos podían poseer y hacer uso de una libertad completa e irrestricta para imprimir aquello que no esté expresamente prohibido” (artículo 12).²⁴²

La transparencia inicia o se enfoca más hacia los actos públicos, de gobierno. Y es lógico esto ya que los recursos públicos mal utilizados y el abuso del poder han sido síntomas oprobiosos que generan corrupción e impunidad. Sin embargo, también en el sector privado se ha construido una cultura de transparencia al interior de cada empresa, que de cierta manera va de la mano con el hecho de demostrar que sus actos no están relacionados con hechos de corrupción, lavado de dinero o alguna actividad ilícita. Entre los temas que integran la política de transparencia de un gobierno destacan: gasto público o presupuesto, ingresos, relaciones con ciudadanos y empresarios, contrataciones, obras, salarios y adquisiciones. Por otro lado, la transparencia en el sector privado evita la corrupción, reduce la evasión de impuestos, protege a los trabajadores, fomenta la competencia económica, da certeza y tranquilidad a los *stakeholders* y genera indicadores para la toma de decisiones, sobre todo en el sector financiero y de inversiones bursátiles.

Hay que decir que, si bien la transparencia se ha desarrollado de manera amplia en muchas actividades del hombre, el sector tecnológico y más específicamente, el sector que utiliza ML, está apenas construyendo su propio sistema de transparencia. Existe una imagen de manipulación de la información que

already been or may in future be issued from the Office of the Chancellor of Justice; although no one may be obliged to obtain and print more of all this, either *in extenso* or abridged as a *species facti*, than he himself requests and regards as adequate and which, when requested, shall immediately be issued to anyone who applies for them, on penalty of the provisions in the following paragraph; but in criminal cases that have been settled by an amicable reconciliation between private individuals no one may, without the agreement of the parties, make use of this freedom as long as they remain alive; while also, if anything concerning grave and unfamiliar misdeeds and abominations, blasphemies against God and the Head of State, evil and cunning schemes in these and other serious criminal cases, superstitions and other such matters should appear in court proceedings or judgments, they shall be completely excluded.” https://www.chydenius.net/tiedostot/worlds_first_foia.pdf [20 de julio de 2022]

²⁴² *Idem*.

está relacionada con el uso de algoritmos de ML, tanto en el sector público como en el sector privado, y esta manipulación o, por lo menos, de una transparencia disminuida deriva de razones reales o personales, de razones legales o de falta de razones (por intereses personales o de grupo).

Existen *razones reales o personales* cuando como usuarios decidimos que nadie entre a nuestros datos o que no compartiremos nuestros datos para determinados fines. Seleccionamos la negativa o simplemente ponemos una contraseña en nuestra cuenta. Las razones legales obligan a quienes tienen conocimiento de determinada información a mantenerla en secreto; por ejemplo, un empleado bancario está obligado tanto por la autoridad legal como por las condiciones de empleo a no revelar los saldos de los clientes a sus amigos. Y, la falta de razones la encontramos en aquellos intentos deliberados de ocultar la información por una decisión personal o de grupo ante el temor de que la revelación de la información genere algún daño y se evita darla o se manipula la entrega. Por ejemplo, una empresa podría responder a una solicitud de información entregando 15 mil páginas de documentos, mal testados y desordenados, lo cual obligaría al solicitante a invertir mucho tiempo buscando la información requerida.²⁴³

La creación de un algoritmo y, principalmente de los algoritmos que se utilizan sin la participación directa de un ser humano generan problemas de transparencia. Existen dos tipos de modelos algorítmicos: Uno abierto y claramente interpretable y explicable; y otro, que se considera de caja negra, porque se conocen los datos de entrada y los de salida, pero no cómo se generaron. No se sabe cómo se entrenan ni cómo llegan a la mejor solución o conclusión. Aquí la transparencia de pronto se ve afectada. Las llamadas para mejorar la transparencia de estos sistemas se encuentran también con limitaciones sociales, legales y técnicas que ponen a la reflexión si la transparencia es la solución para detectar o no la responsabilidad algorítmica.

Entendamos el contexto partiendo del concepto responsabilidad en una sociedad informatizada. Hace más de 25 años, Helen Nissenbaum, hizo una

²⁴³ Frank Pasquale, *The Black Box... op. cit.*, p. 6

reflexión muy novedosa respecto del concepto responsabilidad en una sociedad informatizada. Señaló cuatro tendencias que diluyen las líneas de responsabilidad: “Primero, los sistemas computarizados generalmente son diseñados y operados por “muchas manos”. Como resultado, la responsabilidad se difunde entre muchas partes y se oscurece un análisis de quién o qué instituciones pueden rendir cuentas. En segundo lugar, la inevitabilidad de los errores en el software, lo suficientemente malo como es, también parece proporcionar una especie de excusa para dejar de preocuparse por un software defectuoso en general. Esto tiende a ocultar un análisis adecuado de los errores en el software y dificulta la inversión de esfuerzos en su eliminación y prevención. En tercer lugar, cuando las cosas salen mal, con demasiada facilidad tendemos a culpar a la computadora en lugar de a los humanos involucrados. Finalmente, no es útil que los desarrolladores de software se nieguen rotundamente a rendir cuentas por sus productos.”²⁴⁴

Ibo van de Poel comenta que quizás fue Denise Thompson, la primera en utilizar la noción *The many hands problem* (muchas manos) en un artículo sobre la responsabilidad de los funcionarios públicos y lo describe de la siguiente manera: “Debido a que muchos funcionarios diferentes contribuyen de muchas maneras a las decisiones y políticas del gobierno, es difícil, incluso en principio, identificar quién es moralmente responsable de los resultados políticos”²⁴⁵

A estas cuatro tendencias que diluyen la responsabilidad, debemos adicionar ahora una quinta tendencia que tiene que ver con la creciente importancia de la toma de decisiones, respaldada por el uso de algoritmos de ML. El aprendizaje automático es la herramienta con la que se crean algoritmos para una amplia gama de usos. Tales algoritmos están guiando nuestras vidas. Desafortunadamente, estamos viendo que algunos de sus resultados -como otras obras humanas- traen

²⁴⁴ Cfr. Helen Nissenbaum, “Accountability in a Computerized Society”, en *Science and Engineering Ethics*, número 2, vol. 25, 1996. Consúltese en: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.77> [20 de julio de 2022]

²⁴⁵ Cfr. van de Poel, I., Fahlquist, J. N., Doorn, N., Zwart, S., & Royakkers, L. (2012). The problem of many hands: climate change as an example. *Science and engineering ethics*, 18(1), 49–67, p. 50. Consúltese en: <https://doi.org/10.1007/s11948-011-9276-0> [20 de julio de 2022] Para profundizar más sobre este tema puede consultarse: a Ibo van de Poel, Lambèr Royakkers, Sjoerd D. Zwart, *Moral Responsibility and the Problem of Many Hands*, Routledge, 2018.

sesgos para determinados individuos, grupos o la sociedad. A nivel colectivo, los resultados pueden estar sesgados en función del género, la raza u otras características o circunstancias.

Paul B. de Laat, filósofo experto en estos temas, considera que todo esto se suma a la opacidad como un obstáculo adicional para la responsabilidad en relación con los sistemas de IA basada en Machine Learning – Deep Learning (sobre todo, aquellos que involucran toma de decisiones algorítmicas). “Mientras reina la opacidad, la discusión sobre la toma de decisiones algorítmicas difícilmente puede comenzar. Sin embargo, se pueden inferir algunas características del algoritmo involucrado: se pueden auditar mediante Ingeniería inversa. Desde el exterior, la caja negra se prueba, ya sea insertando varias entradas y observando las salidas, o si la entrada no se puede manipular, observando las salidas disponibles. Sin embargo, esta auditoría requiere mucha mano de obra y solo puede producir información limitada sobre el algoritmo.”²⁴⁶ Al final, tampoco se convierte en una solución. ¿Cómo crear una cultura de responsabilidad algorítmica?

El camino obvio es la transparencia y su relevancia está ganando adeptos en muchos sectores, tanto en los Estados Unidos como en Europa. Laat afirma que la responsabilidad algorítmica hoy se ve desde una perspectiva moral y cita a Zarsky promotor de la transparencia algorítmica: “El análisis del nexo entre transparencia y rendición de cuentas solo puede proceder después de la introducción de dos distinciones importantes. Por un lado, se deben distinguir varias fases de la toma de decisiones: Primero es la recopilación de datos (conjuntos), posteriormente los datos se utilizan en el aprendizaje automático para desarrollar un modelo (desarrollo del modelo), y finalmente ese modelo se utiliza para la toma de decisiones (uso del modelo)”.²⁴⁷

²⁴⁶ Laat, P.B., “Algorithmic Decision-Making Based on *Machine Learning* from Big Data: Can Transparency Restore Accountability?”, *Philos. Technol.*, número 31, 2018, pp. 525–541. <https://doi.org/10.1007/s13347-017-0293-z> [20 de julio de 2022]

²⁴⁷ T. Z., Zarsky, “Transparent predictions”, *University of Illinois Law Review*, número 4, 2013, pp. 1503-1570.

En conjunto, estas dos distinciones implican que cualquier llamado a la transparencia debe ser específico sobre lo que se debe revelar y a quién se debe dar. El aprendizaje automático está muy relacionado entre sí en las etapas de recopilación de datos, construcción del modelo y uso del modelo. Cuando se llama a una organización a rendir cuentas se debe analizar todo para no afectarla.

La posición que apoya Laat es la transparencia de todo el ciclo algorítmico, pero de forma moderada ya que coincide con los diversos contraargumentos expuestos. Primero, en cuanto a la privacidad, afirma que no sería prudente hacer que los conjuntos de datos subyacentes estén disponibles gratuitamente para cualquiera; equivaldría a una invitación por violaciones de la privacidad. En segundo lugar, la total transparencia con respecto a los modelos de aprendizaje automático en uso puede invitar a los interesados a jugar con el sistema y, por lo tanto, socavar su eficiencia. Es importante, por tanto, excluir al público en general de obtener una transparencia total. La misma conclusión se aplica a los modelos que pueden provocar estigmatización; restringir su divulgación a los órganos de supervisión solo parece indicado. En tercer lugar, como regla general, las empresas hacen énfasis en sus derechos de propiedad sobre algoritmos. Esta postura debe evaluarse bien ya que la única alternativa viable para completar la opacidad es, nuevamente, la transparencia limitada a las partes intermedias involucradas en la supervisión.²⁴⁸

Una segunda conclusión de Laat es que, en conjunto, las personas afectadas deben tener derecho a obtener una explicación completa sobre las decisiones que les conciernen. Los organismos de supervisión pueden ayudar a garantizar que esas cuentas se adquieran. La capacidad de interpretación de los algoritmos sigue siendo aquí un tema de la mayor relevancia. ¿Nos mantenemos en la trayectoria dominante de desarrollar algoritmos que se vuelven cada vez más precisos pero cada vez menos inteligibles? ¿U optamos en cambio por otra trayectoria que requiera explicaciones para las decisiones, que se puede lograr ya sea haciendo que los resultados del modelo sean comprensibles ex post o eligiendo modelos ex

²⁴⁸ P.B. Laat, "Algorithmic Decision-Making Based on *Machine Learning* from Big Data: Can Transparency Restore Accountability?" *op. cit.*, pp. 525–541 [20 de julio de 2022]

inocentes también pueden quedar al descubierto. Una sociedad completamente transparente sería una pesadilla de invasión de privacidad... A veces, el camino hacia una investigación ordenada y productiva es confiar el trabajo a un pequeño grupo de expertos... A esta tendencia general la llamo “transparencia calificada”: limitar las revelaciones para respetar todos los intereses involucrados en una determinada información.”.²⁵⁰

La transparencia está implícita en la concepción más básica de la rendición de cuentas. Si no podemos saber lo que está haciendo un gobierno, nuestra empresa o los que hacen otros con nuestros datos, no se puede responsabilizar a alguien de un uso inadecuado. Por ello, contar con algoritmos transparentes o con sistemas de inteligencia artificial transparentes va más allá de esa obvia razón. Se piden sistemas algorítmicos transparentes porque se están volviendo muy importantes para nuestras vidas y para la economía; y, sin embargo, algunos de ellos utilizan modelos y algoritmos cuyo funcionamiento es demasiado complejo para que los siga la mente humana. En el *Panel for the Future of Science and Technology* se construyeron siete áreas de la IA basada en Machine Learning – Deep Learning sobre las que se podía exigir transparencia:

1. Datos. La transparencia de los datos utilizados por el sistema algorítmico, en particular por los algoritmos de aprendizaje automático y aprendizaje profundo, puede referirse a los datos sin procesar, a las fuentes de los datos, a cómo se preprocesaron los datos, a los métodos mediante los cuales se verificaron como imparcial y representativo (incluida la búsqueda de características que sean sustitutos de información sobre clases protegidas), o de los procesos mediante los cuales los datos se actualizan y el sistema se vuelve a entrenar sobre ellos.

2. Algoritmos. La transparencia de los algoritmos de los sistemas puede referirse a probar su salida con entradas para las que conocemos la salida adecuada, reduciendo las variables a las más significativas para que podamos validarlas, **probando el sistema con contrafactuales**²⁵¹ para ver si datos perjudiciales están afectando la salida, una revisión de código de terceros, análisis de cómo funcionan los algoritmos, inspección de informes de errores internos y externos o garantía de que los procesos de desarrollo de software son sólidos.

²⁵⁰ Pasquale, F. *The Black Box Society*... *op. cit.*, pp. 141-143.

²⁵¹ Los contrafactuales describen una dependencia de los hechos externos que conducen a esa decisión sin la necesidad de transmitir el estado interno o la lógica de un algoritmo como resultado. Para mayor detalle de los contrafactuales en el ML *consúltese a:* Wachter, Sandra, Mittelstadt, Brent y Russell, Chris, “Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR” *Harvard Journal of Law & Technology*, volume 31, número 2, 2018. Consúltese en: <https://ssrn.com/abstract=3063289> or <http://dx.doi.org/10.2139/ssrn.3063289> [20 de julio de 2022]

3. Objetivos. Los sistemas algorítmicos también pueden ser transparentes sobre sus objetivos. Cuando un sistema tiene múltiples objetivos, esto significaría ser transparente sobre sus prioridades relativas. Por ejemplo, la IA que conduce vehículos autónomos (AV) podría tener como objetivo reducir las muertes por accidentes de tráfico, disminuir el impacto ambiental de los AV, reducir las lesiones graves, acortar los tiempos de tránsito, evitar daños a la propiedad y proporcionar un viaje cómodo. Se podría requerir que un fabricante sea transparente sobre esos objetivos y su prioridad.

4. Resultados. Se podría exigir a los fabricantes u operadores que sean transparentes sobre los resultados del despliegue de sus sistemas algorítmicos, incluidos los estados internos del sistema (¿qué tan gastados están los frenos de un AV? ¿Cuánta electricidad se consume?), Los efectos en los sistemas externos (¿cuántos accidentes, o veces ha provocado que otro AV se desvíe?), e interacciones basadas en computadora con otros sistemas algorítmicos (¿qué comunicaciones con otros AV, qué datos ingresaron a los sistemas de monitoreo de tráfico?)

5. Cumplimiento. Se puede exigir a los fabricantes u operadores que sean transparentes sobre su cumplimiento general de los requisitos de transparencia que se les hayan impuesto [...]

6. Influencia. Así como el público tiene interés en saber si una parte interesada pagó un artículo en un periódico, el público puede tener interés en saber si algún elemento del proceso de IA se inclinó intencionalmente para favorecer un resultado en particular. Por ejemplo, si una plataforma de búsqueda confiable está impulsando artificialmente algunos resultados porque se les pagó, y si no señala ese hecho a los usuarios, los usuarios pueden ser manipulados. Es posible que los reguladores quieran insistir en que se reconozca de manera notoria dicha influencia.

7. Uso. Los usuarios pueden querer saber qué datos personales está usando un sistema, ya sea para personalizar los resultados o como datos que puedan entrenar al sistema para refinarlo o actualizarlo. Al saber qué datos personales se utilizan, es posible que deseen controlar ese uso, tal vez para que sus resultados personalizados sean más precisos o, más urgentemente, porque sienten que el uso viola su privacidad [...] Aquí también hay áreas grises: recopilar información anónima y muy detallada sobre los viajes realizados por vehículos autónomos (la frecuencia con la que el automóvil frena o se desvía, por ejemplo) podría ser importante para optimizar el tráfico para la seguridad o la eficiencia del combustible.”²⁵²

Se debe tener en cuenta que la transparencia tiene diferentes significados en esta categorización. Puede significar: 1) acceso, previa solicitud, al público o personas autorizadas; 2) Hacer la información pública; 3) inspección directa de procesos internos; 4) resultados de las pruebas del fabricante o del operador del sistema para verificar su precisión y equidad; 5) entrega de subsistemas completos

²⁵² Cfr. Panel for the Future of Science and Technology, *A governance framework for algorithmic accountability and transparency*, European Parliamentary Research Service of the Secretariat of the European Parliament. Bruselas, 2019. Consúltase en: [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU\(2019\)624262_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf) [20 de julio de 2022]

y sus datos para pruebas por personas autorizadas, con los resultados informados al público o a los organismos reguladores; y, 6) acceso a informáticos y administradores para explicar procesos algorítmicos.

Por otro lado, se deben comentar algunas ideas respecto a la interfaz de usuario. Es decir, el lugar en el que debe ocurrir la transparencia algorítmica y el reto es lograr que la interfaz permita al sistema ser transparente *per se* ante el usuario. “Las interfaces de usuario son el punto de encuentro entre el usuario y un sistema computarizado y sus algoritmos y, para muchos usuarios, representan el sistema y sus capacidades. Los algoritmos informáticos y los análisis juegan un papel cada vez más importante en la vida de los ciudadanos, ya que subyacen a los servicios de información populares y las tecnologías “inteligentes”, que se están adoptando rápidamente en todos los sectores de la sociedad, desde los servicios financieros hasta el transporte, la educación y la atención médica.” ²⁵³

Los algoritmos permiten la explotación de fuentes de datos ricas y variadas en cuanto a la información que se puede obtener, para apoyar la toma de decisiones humanas y/o tomar acciones directas. Sin embargo, existen crecientes preocupaciones en torno a su transparencia. Cada vez se reconoce más que, incluso cuando los diseñadores e ingenieros tienen las mejores intenciones, los sistemas que dependen de procesos algorítmicos pueden dar como resultado consecuencias dañinas en el mundo social, como sesgos en sus resultados que pueden dar lugar a la discriminación contra individuos y / o grupos de personas.

Además, muchas personas desconocen por completo los algoritmos que median sus interacciones, incluso en los sistemas que usan a diario. Y aunque lo supieran, es posible que no entiendan que sus propias acciones y elecciones retroalimentan el comportamiento de los algoritmos, que continuamente intentan explotar los rastros digitales de los usuarios para mejorar el rendimiento y / o adaptar la llamada “experiencia del usuario”. La sofisticación de las habilidades digitales que

²⁵³ Styliani Kleanthous, Tsvi Kuflik, Jahna Otterbacher, Alan Hartman, Casey Dugan, and Veronika Bogina, “Intelligent user interfaces for algorithmic transparency in emerging technologies”, en *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion (IUI '19)*, Association for Computing Machinery, New York, 2019, pp. 129.130. DOI: <https://doi.org/10.1145/3308557.3313125> [20 de julio de 2022]

los usuarios requieren para ser consumidores críticos de información, totalmente comprometidos con nuestra sociedad basada en datos, aumenta constantemente.”²⁵⁴

Los expertos en temas algorítmicos coinciden en que muchos procesos son opacos para el usuario, porque son demasiado complejos técnicamente para ser explicados y/o son secretos comerciales protegidos. Por ejemplo, a menudo es difícil interpretar los resultados. a partir de modelos inducidos por nuevas técnicas de aprendizaje automático, como el aprendizaje profundo, que plantean importantes desafíos técnicos para la transparencia. Dadas estas consideraciones, la comunidad necesita centrarse en las herramientas del usuario (tanto técnicas como educativas) para ayudarles a tomar conciencia de los procesos algorítmicos en los sistemas que utilizan.

En mayo de 2017, la *Association for Computing Machinery* (ACM) aprobó una declaración sobre “Transparencia algorítmica y responsabilidad”. La declaración de ACM presenta un conjunto de principios para promover la transparencia algorítmica.²⁵⁵

El objetivo de IUI ATEC es enfocarse en tres de los siete principios que describen los enfoques para combatir los sesgos algorítmicos que pueden ser aplicados por los investigadores, incluso sin acceso a las interacciones de un sistema dado. Se hizo un taller cuyo objeto fue reunir una participación académica y de la industria para atraer la atención de una audiencia interdisciplinaria que incluye investigadores y profesionales que trabajan en los campos de IUI, sistemas,

²⁵⁴ *Idem*.

²⁵⁵ Cfr. Association for Computing Machinery, *Statement on Algorithmic Transparency and Accountability* Public Policy Council, aprobada el 12 de enero de 2017. Consúltase en: http://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf Consúltase también: Leslie, David, *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. 2019. <https://doi.org/10.5281/zenodo.3240529> y Guía Marco de Ética de datos Orientación para las organizaciones del sector público sobre cómo utilizar los datos de manera adecuada y responsable al planificar, implementar y evaluar una nueva política o servicio: Central Digital and Data Office, *Guidance for public sector organisations on how to use data appropriately and responsibly when planning, implementing, and evaluating a new policy or service*, 16 de septiembre de 2020. Consúltase en: <https://www.gov.uk/government/publications/data-ethics-framework> [Actualización a 20 de julio de 2022]

experiencia del usuario y herramientas centradas en el usuario para promover TA (*Cyprus Center of Algorithmic; University of Haifa; IBM*). Las presentaciones hechas por los participantes cubrieron los siguientes temas:

- a. transparencia del sistema de asesoramiento
- b. transparencia algorítmica de los agentes de conversación
- c. dimensiones de transparencia (auditabilidad y explicabilidad)
- d. transparencia de las recomendaciones educativas
- e. transparencia en la realidad aumentada

Los principios para la rendición de cuentas algorítmica de la mencionada declaración conjunta son los siguientes:

Conciencia: los propietarios, diseñadores, constructores, usuarios y otras partes interesadas de los sistemas analíticos deben ser conscientes de los posibles sesgos involucrados en su diseño, implementación y uso, y el daño potencial que los sesgos pueden causar a las personas y a la sociedad.

Acceso y compensación: los reguladores deben alentar la adopción de mecanismos que permitan el interrogatorio y la reparación para individuos y grupos que se ven afectados negativamente por decisiones algorítmicamente informadas.

Responsabilidad: las instituciones deben ser responsables de las decisiones tomadas por los algoritmos que utilizan, incluso si no es factible explicar en detalle cómo los algoritmos producen sus resultados.

Explicabilidad: Se alienta a los sistemas e instituciones que utilizan la toma de decisiones algorítmicas a que produzcan explicaciones sobre los procedimientos seguidos por el algoritmo y las decisiones específicas que se toman. Esto es particularmente importante en contextos de políticas públicas.

Procedencia de los datos: los constructores de los algoritmos deben mantener una descripción de la forma en que se recopilaban los datos de entrenamiento, acompañados de una exploración de los posibles sesgos inducidos por el proceso humano o algorítmico de recopilación de datos. El escrutinio público de los datos ofrece la máxima oportunidad de correcciones. Sin embargo, las preocupaciones sobre la privacidad, la protección de los secretos comerciales o la revelación de análisis que podrían permitir a los actores maliciosos jugar con el sistema pueden justificar la restricción del acceso a personas calificadas y autorizadas.

Auditabilidad: los modelos, algoritmos, datos y decisiones deben registrarse para que puedan auditarse en los casos en que se sospeche un daño.

Validación y prueba: las instituciones deben utilizar métodos rigurosos para validar sus modelos y documentar esos métodos y resultados. En particular, deben realizar pruebas

de rutina para evaluar y determinar si el modelo genera daños discriminatorios. Se alienta a las instituciones a hacer públicos los resultados de tales pruebas.²⁵⁶

René F. Kizilcec comenta que la creciente prevalencia de las interfaces algorítmicas plantea preguntas nuevas para diseñadores, académicos y críticos. Por tanto, la investigación se centra en cómo el diseño transparente de interfaces algorítmicas puede promover la conciencia y fomentar la confianza. La hipótesis se basa en cómo la transparencia afecta la confianza basándose en las teorías del procesamiento de la información y la justicia procesal. Se elaboró un experimento para probar tres niveles de transparencia del sistema. Las personas cuyas expectativas se violaron (al recibir una calificación inferior a la esperada) confiaron menos en el sistema que el algoritmo de calificación. Proporcionar demasiada información erosionó esta confianza. Las actitudes de las personas cuyas expectativas se cumplieron no variaron con la transparencia. Los resultados se

²⁵⁶ The Principles for Algorithmic Transparency and Accountability from the joint statement are as follows:

1. Awareness: Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.
2. Access and redress: Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.
3. Accountability: Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.
4. Explanation: Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.
5. Data Provenance: A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.
6. Auditability: Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.
7. Validation and Testing: Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

Cfr. Association for Computing Machinery, *Statement on Algorithmic Transparency and Accountability Public Policy Council*, ACM Europe Policy Committee, aprobada el 12 de enero de 2017. Consúltase en: http://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf [20 de julio de 2022]

discuten en términos de un modelo de proceso dual de cambio de actitud y la profundidad de la justificación de la inconsistencia percibida.²⁵⁷

Los avances de la IA basada en Machine Learning – Deep Learning apuntan a enfrentar el desafío de administrar una gran cantidad de información para el consumo humano. Las interfaces algorítmicas que seleccionan las noticias en línea crean estaciones de radio personalizadas y personalizan los resultados de búsqueda se han convertido en algo común y aparentemente indispensable. Sin embargo, muchas personas desconocen la inteligencia oculta de estos sistemas a pesar de su impacto potencial en la sociedad. Esto plantea un nuevo conjunto de preguntas para diseñadores, académicos y críticos de los medios.

La transparencia puede promover o erosionar la confianza de los usuarios en un sistema al cambiar las creencias sobre su confiabilidad. La confianza es una preocupación clave en el diseño de la tecnología, ya que afecta la adopción inicial y el uso continuo de las tecnologías. Teniendo en cuenta la tendencia de las personas a tratar las nuevas tecnologías como actores sociales, la definición actual de confianza se basa en el trabajo previo en contextos interpersonales fuera de línea. La confianza se entiende como “una actitud de expectativa segura en una situación en línea de riesgo de que las vulnerabilidades de uno no sean explotadas”. Una forma de asegurar a las personas que no serán explotadas es a través de la transparencia en el diseño, que puede fomentar una mejor comprensión del sistema y la medida en que es justo y preciso.²⁵⁸

Joy Lu, Dokyun, Lee, Tae Wan Kim, and David Danks comentan que los algoritmos de aprendizaje automático han ganado un uso generalizado, tanto para proporcionar predicciones a usuarios expertos como para recomendar decisiones a usuarios cotidianos. Sin embargo, estos sistemas de IA a menudo son cajas negras, los usuarios finales rara vez reciben una explicación. La necesidad crítica de la

²⁵⁷ René F., Kizilcec, “How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. Proceedings of the 2016 CHI”, *Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, 2016, pp. 2390–2395. DOI: <https://doi.org/10.1145/2858036.2858402>

²⁵⁸ *Idem*.

explicación por parte de los sistemas de IA ha llevado a reclamos de transparencia algorítmica, incluido el derecho a la explicación del Reglamento General de Protección de Datos de la Unión Europea.²⁵⁹

Estos autores concluyeron que es necesario construir 1) un marco generalizable basado en Filosofía, psicología y aprendizaje automático interpretable para investigar y definir características de buena explicación, y 2) realizar un experimento de laboratorio a gran escala para medir el impacto de diferentes factores en percepciones de la gente de comprensión, intención de uso y confianza en los sistemas de IA.²⁶⁰

La investigación también proporciona una guía para la alta gerencia sobre cómo presentar los fundamentos de predicción algorítmica a los usuarios finales para fomentar la confianza y la adopción, y elementos de explicación y transparencia que deben ser considerados por los investigadores e ingenieros de IA al diseñar, desarrollar e implementar transparentes o algoritmos explicables. Según una encuesta de IBM, alrededor del 60% de los 5.000 ejecutivos estaban preocupados “por poder explicar cómo AI está utilizando los datos y tomar decisiones”²⁶¹. Joy Lu, Dokyun Lee, Tae Wan Kim y David Danks comparten preocupaciones como la de crear una Declaración de Derechos Algorítmica destinada a proteger a los consumidores de las consecuencias nocivas e involuntarias de la IA. Sin embargo, los algoritmos más exitosos hoy en día no son transparentes.²⁶²

Se ha demostrado que las diferentes variedades de modelos de redes neuronales profundas y los modelos de árboles de decisiones a menudo reclaman la primera posición en competencias de ciencia de datos. Todos estos modelos son fundamentalmente “cajas negras” que incluyen muchas capas de entradas

²⁵⁹ Joy Lu, Dokyun, Lee, Tae Wan Kim, and David Danks. “Good Explanation for Algorithmic Transparency. Proceedings of the AAAI/ACM” *Conference on AI, Ethics, and Society*, Association for Computing Machinery, New York, número 93, 2020 en DOI: <https://doi.org/10.1145/3375627.3375821>

²⁶⁰ *Idem.*

²⁶¹ *Idem.*

²⁶² *Idem.*

complejas, típicamente no lineales. Puede ser bastante difícil para cualquiera comprender la salida del algoritmo y / o por qué el modelo hace predicciones clave. Dados estos desafíos, los esfuerzos para desarrollar algoritmos más interpretables, explicables o inteligibles comprenden un área clave de la investigación actual.²⁶³

La necesidad de explicaciones por parte de los sistemas de IA ha llevado a la regulación de la transparencia algorítmica. Los autores ubican ejemplos como, el Reglamento General de Protección de Datos de la UE (GDPR) promulgado en 2018, el cual requiere que las compañías brinden una explicación ex post a las partes involucradas (aunque con poca claridad sobre lo que se considera “significativo”). La Ley de Igualdad de Oportunidades de Crédito exige que las compañías financieras brinden razones para tomar decisiones a sus clientes, como por qué se denegó un préstamo. También, un número creciente de investigadores ha estado desarrollando IA *explicable eXplainable* (XAI).

Respecto a la privacidad. El derecho a la privacidad conjuntamente con la seguridad de un Estado son algunos de los límites que restringen la transparencia. El derecho a la privacidad se expone como tema de reflexión jurídica el 15 de diciembre de 1890 en la *Harvard Law Review*. Samuel D. Warren y Louis D. Brandeis afirman que es un derecho consuetudinario inviolable que posee cada individuo al determinar, ordinariamente, “en qué medida sus pensamientos, sentimientos y emociones deben ser comunicados a otros... fijando los límites de la publicidad que se les dará.”²⁶⁴ Para estos autores, el derecho a la privacidad significa que cada individuo tenía derecho a elegir compartir o no compartir con otros la información. sobre su “vida privada, hábitos, actos y relaciones”.

Warren y Brandeis argumentaron que era necesario que el sistema legal reconociera el derecho a la privacidad porque cuando la información sobre la vida privada de un individuo se pone a disposición de otros, tiende a influir e incluso a dañar el núcleo mismo de la personalidad de un individuo: “su estima de sí mismo”.

²⁶³ *Idem*.

²⁶⁴ Warren y Brandeis, “The Right to Privacy”, *Harvard Review*, volumen 4, número 5, diciembre 15, 1890. <https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf>

El concepto original de Warren y Brandeis sobre el derecho a la privacidad encarnaba así en una “percepción psicológica, en ese momento inexplorada, de que la personalidad de un individuo, especialmente su propia imagen, puede verse afectada, y a veces distorsionada o dañada, cuando la información sobre la vida privada de ese individuo se pone a disposición de otras personas”.²⁶⁵

Este derecho a la privacidad no era nuevo. Warren y Brandeis acuñan su frase “right to privacy” y la de “the right to be let alone” haciendo una revisión histórica de tal reconocimiento y aplicación legal que se encontraba presente en una amplia variedad de precedentes de diferentes áreas del Derecho norteamericano (como contratos, propiedad, fideicomisos, derechos de autor, protección de secretos comerciales y cuestiones procesales) y crean una nueva teoría jurídica que dio forma a este principio.

Según Warren y Brandeis, el derecho a la libertad asegura amplios privilegios civiles, pero no la privacidad. También contrastaron el derecho a la propiedad, que comprendía los intereses materiales del individuo, toda forma de posesión, tanto intangible como tangible, con la preocupación del derecho a la privacidad por los intereses espirituales. Por ejemplo, Thomas Cooley en su *Treatise on the law torts* (1879) parece haber acuñado la frase “el derecho a ser dejado solo”.²⁶⁶

Vittorio Frosini comenta que en *The Right to Privacy* se negaba el derecho a difundir noticias sobre la vida privada de una persona (Warren) para informar a los lectores de los periódicos de Boston y al mismo tiempo se contraponía a la exigencia de dar a conocer a la opinión pública la información, como un derecho fundamental en las democracias. “El derecho a la información presentó así dos aspectos: de una parte, el derecho de informar, o sea, de difundir noticias; de otra parte, el derecho de ser informado, o sea de exigir noticias... [encontrando] un límite en el derecho a la intimidad; pero él se configura, además, como un nuevo derecho público

²⁶⁵ Ibidem.

²⁶⁶ Existe una explicación amplia sobre el derecho a la privacidad en: Dorothy J. Glancy, “The Invention of the Right to Privacy”, *Arizona Law Review*, volumen 21, número 1, 1979, pp. 1-39. <https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?referer=https://en.wikipedia.org/&httpsredir=1&article=1318&context=facpubs>

subjetivo, que es la libertad de información.”.²⁶⁷ Y añade que, el 17 de octubre de 1970, el Parlamento de la región de Hesse (Alemania) aprobó una ley que creaba el “comisario de protección de las informaciones recogidas en bancos de datos”. Más tarde, el 15 de noviembre de 1976, se aprobaría la Ley Federal sobre la Protección de las Informaciones... en la cual queda instituida la figura del comisionado federal (dentro de la estructura del gobierno) y un responsable de la información dentro de la empresa.²⁶⁸

Para el 28 de enero de 1981, en Estrasburgo, Francia, es adoptado el Convenio para la Protección de las Personas con Respecto al Tratamiento Automatizado de Datos de Carácter Personal²⁶⁹, con la intención de dar soporte legal al problema de la protección de datos personales y la confidencialidad de la información contenida en las bases de datos. Tommaso Edoardo Frosini, hijo de Vittorio Frosini, llega a la conclusión de que la “doctrina de la libertad informática” nació con este documento al dar soporte legal al problema relacionado con “la protección de la confidencialidad con referencia a las bases de datos” y añade:

“...el derecho a la libertad informática manifiesta un nuevo aspecto de la antigua idea de libertad personal y constituye el avance de una nueva frontera de la libertad humana hacia la sociedad futura. Esta nueva forma de libertad personal deriva de la necesidad de salvaguardar a la persona humana de la amenaza y la trampa que representa degeneración del nuevo poder social, económico y legal, que es el poder informático; y por tanto, la capacidad de acumular, memorizar, procesar y transmitir datos informáticos personales, lo que le confiere un poder cognitivo previamente desconocido y que permite la implementación de una vigilancia oculta, omnipresente y omnipresente de la conducta privada”. En su versión original, la expuesta en 1981, la libertad informática se representaba, como la libertad política, como positiva y negativa. La libertad informática negativa expresa “el derecho a no hacer pública determinada información de carácter personal, privado y confidencial (estas calificaciones, que en algunos casos pueden no coincidir entre sí)”; la libertad informática positiva, por otro lado, expresa la facultad de “ejercer un derecho de control sobre los datos personales que han escapado del círculo de la privacidad

²⁶⁷ Vittorio, Frosini, *Cibernética, Derecho, Internet y Sociedad*, Lex. Ediciones Olejnik, Santiago, Anexo II, pp. 116 y 117.

²⁶⁸ *Idem*.

²⁶⁹ En el Diario Oficial de la Federación del 28 de septiembre de 2018 se publicó el *Decreto Promulgatorio del Convenio para la Protección de las Personas con respecto al Tratamiento Automatizado de Datos de Carácter Personal*. El Convenio mencionado fue aprobado por la Cámara de Senadores del Congreso de la Unión, el 26 de abril de 2018. El instrumento de adhesión fue depositado ante el secretario general del Consejo de Europa, de conformidad con lo establecido en el artículo 23, numeral 2, del Convenio para la Protección de las Personas con Respecto al Tratamiento Automatizado de Datos de Carácter Personal.

para convertirse en elementos de entrada de un programa electrónico; y por tanto la libertad informática positiva, o el derecho subjetivo reconocido, de conocer, corregir, eliminar o incorporar datos en un archivo personal electrónico “.²⁷⁰

Actualmente, las empresas que están en la World Wide Web recopilan cada vez más datos sobre sus usuarios, pero son discretas o muy silenciosas al momento de que se les pide conocer cómo los utilizan, resguardan, clasifican y, sobre todo, explotan comercialmente. Tampoco son muy activas ante las regulaciones que permiten una carga de responsabilidad o un incremento de derechos para sus usuarios ejerciendo un control absoluto sobre los expedientes digitales que resguardan.

A medida que las nuevas tecnologías van avanzando, aumenta la necesidad de contar con datos ya sea para procesar dicha información o para producir más información que permita tomar mejores decisiones. La *data* o los datos son la energía que requieren los sistemas de ML para funcionar y el tesoro más preciado de las organizaciones. Los científicos de datos, una nueva profesión que ha cobrado gran auge en las empresas, son los responsables de emitir un informe analizando los datos para aumentar el conocimiento de la organización e indagar la mejor manera de utilizarlos para darle “valor” a la empresa. Esto último, el generar datos y que a la vez se suban a la web o que utilicemos la web como parte de nuestra vida familiar, laboral o social está construyendo los perfiles digitales de las personas y empresas. Seguramente, las personas que los analizan o los sistemas de IA que los interpretan sabrán más detalles sobre nuestros hábitos, creencias, aficiones, intereses, etcétera.²⁷¹

²⁷⁰ Cfr. *Informatica e diritto*, XXXIV annata, Vol. XVII, 2008, n. 1-2, pp. 87-97 en: [frontespizio e dedica.qxd \(cnr.it\)](http://frontespizio.e dedica.qxd (cnr.it))

²⁷¹ Un ejemplo de empresa exitosa que incrementó exponencialmente sus ganancias y resultados gracias a la explotación de los datos es BANORTE. José Antonio Murillo, director general de Analítica del Grupo Financiero Banorte explica cómo se convirtió de economista a científico de datos y, posteriormente, como a través de un equipo de científicos de datos posicionó al banco por encima de sus competidores. En todo esto encontramos los datos que tiene la organización y su forma de utilización. Es un caso de éxito estudiado ya en varias partes del mundo. Cfr. Colegio de economistas de Oaxaca, “Analítica e Inteligencia Artificial: El caso de éxito de BANORTE”. (11 de febrero de 2021) en: <https://www.youtube.com/watch?v=QZr2UxIJm7o>

Esto trae como consecuencia un descontrol o, por lo menos, disminución de la privacidad personal. Los evaluadores de crédito, los motores de búsqueda, los sistemas financieros toman los datos y los convierten en puntajes, clasificaciones, cálculos de riesgo y listas de vigilancia con consecuencias de vital importancia. Pero los algoritmos mediante los cuales lo hacen de pronto se vuelven inmunes a cualquier pregunta o indagación, excepto en aquellas ocasiones en que un denunciante litiga sus derechos, como ocurrió con el famoso caso del llamado “derecho al olvido” (2014) y otros que han venido surgiendo durante la última década²⁷² como Cambridge Analytica. Carissa Véliz, profesora en la Universidad de

²⁷² Julia Powles y Enrique Chaparro narran de forma clara el caso Google-derecho al olvido: “Mario Costeja González se pasó de 2009 a 2014 luchando por eliminar 18 palabras de los resultados de búsqueda de Google en su nombre. Abanlex, despacho especializado en protección de datos personales tomó su caso. Los avisos habían sido publicados en el diario español La Vanguardia y recientemente digitalizados. Costeja González pidió al diario que los retirara. Cuando eso no tuvo éxito, desafió a Google, y el caso finalmente fue elevado al Tribunal de Justicia de las Comunidades Europeas, el tribunal más alto de Europa. En mayo de 2014, el TJCE falló en contra de Google. Reconoció que cuando ingresamos el nombre de alguien como consulta de búsqueda, se presentan de manera mecanicista momentos dispersos de su vida, con un significado distorsionado por la falta de contexto, construyendo un perfil detallado pero selectivo. Entonces, ¿cuáles son los derechos de las personas con las que se relacionan esos perfiles? ¿Y cuáles son los derechos de quienes buscan información? La pregunta produce una interesante división filosófica. Una posición es que una vez en línea, la información debe permanecer en línea (excepto cuando sea ilegal por difamación, derechos de autor o leyes penales). Este es generalmente el punto de partida de la mayoría de las empresas de Internet, las organizaciones de libertad de expresión y los medios de comunicación estadounidenses; un punto de vista típico para aquellos planteados en el pensamiento de la Primera Enmienda de Estados Unidos. Por otro lado, existen todo tipo de razones para eliminar datos, además de la obligación legal. Uno podría querer eliminar información por razones emocionales, razones éticas o “simplemente porque”, cuando no existe un interés compensatorio. Algunas solicitudes de eliminación recientes aprobadas por Google incluían historiales médicos de pacientes; fotos privadas íntimas; e hilos antiguos en conversaciones grupales privadas que terminaron en línea. Incluyen recordatorios destacados de que una persona fue víctima de violación, agresión u otros actos delictivos; que alguna vez fueron testigos incidentales de una tragedia; que sus allegados - pareja o hijo - fueron asesinados. Las fuentes originales suelen tener muchos años o décadas. Son recordatorios estáticos, no representativos de vidas pasadas, que carecen de la dinámica de la realidad. En el mundo real, la información se sedimenta con el tiempo, lo que brinda a las personas la capacidad de seguir adelante, recordando, pero sin sentirse agobiados por su pasado. Fuera de línea, nos comunicamos de diferentes maneras con diferentes “públicos” y propósitos. Y esta fue la opinión del tribunal de Luxemburgo, que extrajo su justificación de la ley de protección de datos de la UE. El tribunal dictaminó que los datos personales deben eliminarse de los resultados de búsqueda del nombre de una persona cuando estén desactualizados, sean inexactos, inadecuados, irrelevantes o carezcan de propósito, y cuando no exista interés público. La frase “derecho al olvido” se mencionó brevemente en la sentencia, pero los medios de comunicación, Google y los reguladores se apoderaron de ella inmediatamente. Aunque ahora reemplazado por el más exacto “derecho a eliminar de la lista”, el impacto de la etiqueta “derecho al olvido” fue forzar el debate en binarios: olvido versus recuerdo, privacidad versus libertad de expresión, censura versus verdad o historia. Se trata de falsas dicotomías, insuficientemente matizadas para afrontar la realidad de nuestras vidas y las complejidades de la existencia humana. El objetivo de tener derechos contra los

Oxford, ubica este segundo evento como el que nos alertó a todos en el tema de la privacidad:

“Después del escándalo de Cambridge Analytica, y de experimentar casos de humillación pública o robo de identidad de nosotros mismos, ahora entendemos que las consecuencias de la falta de privacidad actual son tan graves como lo eran antes de que apareciera Internet. Políticamente, comprometer nuestra privacidad es más peligroso que nunca. Nunca habíamos acumulado tantos datos personales sobre ciudadanos. Y hemos permitido que la vigilancia crezca en un momento en que los estándares de ciberseguridad son deficientes, las democracias son débiles y los regímenes autoritarios con un don para la piratería están en aumento. La tecnología digital utilizó el manto de la invisibilidad de los datos para erosionar nuestra privacidad. Pero ahora conocemos sus trucos. Podemos recuperar el control de nuestros datos personales. [...] Los bloqueos para esos datos no solo deben ser legales (ya que las leyes cambian y se infringen), sino también técnicos (por ejemplo, mediante cifrado) y prácticos. Por práctico me refiero a dificultar el acceso de las personas a esos datos. Si un diario se mantiene en formato papel en un registro en una ciudad, es accesible para investigadores serios, pero a los actores deshonestos les resulta más difícil acceder a él que si se publicara en línea e indexara en motores de búsqueda. La medida en que algo es accesible importa. Esa es la esencia del derecho de Europa al olvido. Cuando Mario Costeja hizo una búsqueda en Google sobre su nombre en 2009, algunos de los primeros elementos que aparecieron fueron un par de avisos de finales de la década de 1990 en el periódico español La Vanguardia. Los avisos eran sobre la subasta de la casa de Costeja para recuperar sus deudas de seguridad social. Primero se habían publicado en la edición impresa del periódico, que luego se digitalizó. Costeja acudió a la Agencia Española de Protección de Datos para denunciar a La Vanguardia. Argumentó que esos avisos ya no eran relevantes, ya que sus deudas habían sido liquidadas. Tener esa mancha ligada a su nombre estaba dañando su vida personal y profesional. El periódico se había negado a borrar los registros, y la Agencia Española de Protección de Datos estuvo de acuerdo, ya que La Vanguardia había publicado legalmente esos registros públicos. Pero la agencia le pidió a Google que eliminara el enlace al aviso de subasta. Una persona que ha pagado sus deudas no debe cargar con ese peso por el resto de su vida. Google apeló la decisión y el caso terminó en el Tribunal de Justicia de la Unión Europea, que en 2014 falló a favor del derecho al olvido. Los registros de Costeja todavía se pueden encontrar en La Vanguardia, pero esos datos no están indexados en la Búsqueda de Google. Si bien la implementación de ese derecho ha suscitado dudas y críticas, su principio tiene sentido. Es cuestionable que empresas privadas sean los árbitros de si una solicitud para hacer algo menos accesible tiene mérito, incluso si la decisión puede ser apelada y remitida a una Agencia de Protección de Datos. Sin embargo, lo que más importa es que el derecho al olvido nos protege de ser perseguidos por datos personales que están desactualizados, son inexactos, inadecuados, irrelevantes o carecen de propósito, y cuando no hay interés público...”²⁷³

motores de búsqueda no es manipular la memoria o eliminar la información, sino hacerla menos prominente, cuando esté justificado, y combatir los efectos secundarios de este fenómeno singularmente moderno de que la información es instantánea, global y perpetuamente accesible.” Cfr. Julia Powles y Enrique Chaparro, “How Google determined our right to be forgotten”, *The Guardian*, 18 de febrero, 2015. <https://www.theguardian.com/technology/2015/feb/18/the-right-be-forgotten-google-search>. También: Kassam, Ashifa, “Spain’s everyday internet warrior who cut free from Google’s tentacles”, *The Guardian*, 13 de mayo, 2014. <https://www.theguardian.com/technology/2014/may/13/spain-everyman-google-mario-costeja-gonzalez>

²⁷³ Carissa, Véliz, *Privacy is Power*, Transworld, Edición de Kindle, 2020. [La traducción es propia]

El Consejo de Europa ha creado normas estrictas para evitar los riesgos significativos para la privacidad y la protección de datos que plantea el uso cada vez mayor de tecnologías de reconocimiento facial. Está ya demostrado que ciertas aplicaciones de reconocimiento facial en ciertos temas deben prohibirse por completo para evitar la discriminación.

Por ejemplo, su uso para determinar el color de piel de una persona, su religión, creencia, sexo, origen étnico, la edad, la salud o la condición social deben levantar banderas rojas al momento de utilizar el reconocimiento facial. También debe evaluarse su uso en el “reconocimiento de afectos”, es decir, la identificación de emociones o sentimientos, condiciones de salud o nivel de compromiso de los trabajadores, ya que plantean riesgos importantes en campos como el empleo, la salud, el acceso a los seguros y la educación.²⁷⁴

El Consejo de Europa ha sido contundente en estos temas y también se ha manifestado en contra del uso de tecnologías encubiertas de reconocimiento facial en vivo por parte de las fuerzas del orden (aceptable solo si es estrictamente necesario y proporcionado para prevenir riesgos inminentes y sustanciales para la seguridad pública que estén documentados con anticipación). No está permitido que las empresas privadas utilicen el reconocimiento facial en entornos no controlados, como centros comerciales, con fines de marketing o seguridad privada.²⁷⁵

Las decisiones críticas o las decisiones de gran envergadura que están tomando las empresas que hacen análisis de datos, no se toman sobre la base de los datos por sí mismos, existe todo un equipo de analistas que estudian y recomiendan determinada resolución, previo procesamiento algorítmico; es decir, hacen cálculos codificados en un software que genera puntajes.

²⁷⁴ Dirección General de Derechos Humanos y Estado de Derecho del Consejo de Europa, Facial recognition: strict regulation is needed to prevent human rights violations, *News 2020-2021*, 28 de enero de 2021. Consúltase en: <https://www.coe.int/en/web/human-rights-rule-of-law/-/facial-recognition-strict-regulation-is-needed-to-prevent-human-rights-violations>

²⁷⁵ Consúltase el Convenio para la Protección de las Personas con respecto al Tratamiento Automático de Datos Personales (CETS No. 108) en: <https://www.coe.int/en/web/data-protection/convention108-and-protocol>

Esto ha llevado a una exigencia de creación, enseñanza y aplicación del derecho a la privacidad y su relación con otros derechos como el derecho a la información y su ponderación judicial en asuntos de regulación normativa o de cumplimiento normativo que garantice la privacidad de la información de las personas en prácticamente todo el mundo. Es importante tener presente que el mayor volumen de datos se genera entre los Estados Unidos y la Unión Europea, cualquier cambio normativo en estas regiones influirá en prácticamente todo el mundo.

El 6 de marzo de 2021, en el periódico *The New York Times*, su comité editorial publica un artículo sobre la privacidad de datos y su relación con el uso de un enfoque de participación voluntario para frenar los excesos de los *Big Tech*: ¿hasta qué punto este enfoque voluntario sirve o solo forma parte de una buena simulación de buenas intenciones? Es decir, pensemos en que cada persona conectada en internet está generando datos día con día (crea con ello su propio expediente). La mayoría de esos datos se utilizan, en automático, para fines comerciales o publicitarios y aunque los consumidores pueden decidir si aceptan o no compartir sus datos, la realidad es que, el no compartir los datos te lleva al no poder utilizar el servicio en su totalidad.

Las *Big Tech* están influyendo en los intentos de construir legislación para imponer a los consumidores la responsabilidad de autorizar o no autorizar la recopilación de sus datos y se ha generado una discusión entre estas empresas y con los gobiernos respecto de las “disposiciones de exclusión voluntaria en manos de los consumidores” ya que no están funcionando para proteger datos. Los usuarios no están dispuestos a leer las políticas, términos y condiciones que ponen las compañías en la recopilación de datos; otras organizaciones prefieren ocultar sus políticas de recopilación de datos en sitios de difícil acceso para el usuario y

otras más ni siquiera cuentan con un aviso de privacidad o deciden ponerlo solo a determinadas subsidiarias o filiales.²⁷⁶

El comité editorial del New York Times se manifiesta a favor de una legislación federal que regule de manera completa el tema de la privacidad de datos en la web, con disposiciones que permitan que la recopilación de datos personales esté disponible solo con el consentimiento previo de los consumidores y recomienda utilizar el Reglamento General de Protección de Datos de la Unión Europea como guía.²⁷⁷ Seguramente esta tendencia se irá multiplicando en otros países, siendo la mejor opción el que las empresas no condicionen a los usuarios a dar su datos a cambio de mantener el servicio, sobre todo en aquellos datos personales que el usuario decide no compartir.

Ya algunos medios de comunicación lo están tomando en cuenta y han cambiado su política de privacidad. Cuando el Reglamento General de Protección de Datos (RGPD) entró en vigor el 25 de mayo de 2018, varios editores bloquearon todas las compras de anuncios de intercambio abierto en sus páginas europeas. Se llegó a la tendencia de la venta directa de anuncios en vez de la venta por comportamiento y segmentación. Según Jean-Christophe Demarta, vicepresidente ejecutivo de publicidad global en *New York Times International*: “El hecho de que ya no ofrezcamos opciones de segmentación por comportamiento en Europa no parece obstaculizar lo que los anunciantes quieren hacer con nosotros”, dijo: “La deseabilidad de una marca puede ser más fuerte que las capacidades de focalización. No nos ha afectado desde el punto de vista de los ingresos y, por el contrario, nuestro negocio de publicidad digital sigue creciendo a buen ritmo “.²⁷⁸

²⁷⁶ Un libro que describe el panorama sobre los 9 gigantes de la tecnología lo ha escrito Amy Webb: *The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity*, Hachette Book Group, Inc. 2019. Existe la versión en español de Ediciones Culturales Paidós (México, 2021).

²⁷⁷ Cfr. Jean-Christophe Demarta, The Editorial Board. America's Privacy Settings Are Wrong. Section SR, *The New York Times*. New York. Marzo 7, 2021, p. 6.

²⁷⁸ Cfr. Jessica, Davies, “After GDPR, The New York Times cut off ad exchanges in Europe - and kept growing ad revenue”. *DIGIDAY*, enero 16, 2019. <https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/>

Editores como USA Today retiraron toda la publicidad en la Unión Europea para evitar sanciones del RGDP. Los Ángeles Times fue advirtiendo que sus anuncios no estarían disponibles en varios países europeos. Sin embargo, se dieron cuenta de que esto no será sostenible para muchas empresas ya que una buena cantidad de sus ingresos dependen de la publicidad que se genera en el mercado europeo.²⁷⁹

Por otro lado, también está el polémico tema de las “cookies de publicidad conductual” que son el eje el marketing digital y de los estudios de viabilidad comercial, financiera o económica de las empresas que utilizan el internet. Las cookies se usan para crear proyectos comerciales, abrir productos al mercado, conocer clientes y competidores, entre otros. Una vez que alguien visita un sitio que muestra uno de sus anuncios, puede adaptar el mensaje del anuncio para satisfacer los intereses inferidos de la persona. La Declaración 03/2021 sobre el Reglamento de privacidad electrónica del European Data Protection Board, adoptada el 9 de marzo de 2021 está modificando todos los avisos de privacidad de las páginas europeas. La declaración señala lo siguiente:

hacer cumplir el requisito de consentimiento para las cookies y tecnologías similares, y ofrecer a los proveedores de servicios herramientas técnicas que les permitan obtener fácilmente dicho consentimiento (Artículo 84) La necesidad de un enfoque de preservación de la privacidad con respecto a las soluciones de “tómalo o déjalo”, la EDPB considera que la necesidad de obtener un consentimiento otorgado libremente debe evitar que los proveedores de servicios utilicen prácticas desleales como las soluciones de “tómalo o déjalo”, que condicionan el acceso a los servicios y funcionalidades al consentimiento del usuario para el almacenamiento de información o para acceder a ella. a la información ya almacenada en el equipo terminal de un usuario (los llamados “muros de cookies”). La EDPB destaca la necesidad de incluir una disposición explícita en el reglamento de privacidad para consagrar esta prohibición, con el fin de permitir a los usuarios aceptar o rechazar la elaboración de perfiles. Por lo tanto, se debe proponer a los usuarios alternativas justas ofrecidas por los mismos proveedores de servicios. Dichos principios deben aplicarse por igual a todos los proveedores de servicios, independientemente de su sector de actividad o de su modelo de financiación actual (véase el considerando 21 bis de la posición del Consejo).²⁸⁰

²⁷⁹ Cfr. Jessica, Davies, “Axel Springer counters Google with its own consent management tool”, *DIGIDAY*. Mayo 30, 2018. <https://digiday.com/media/axel-springer-counters-google-consent-management-tool/>

²⁸⁰ Cfr. EDPB-EDPS Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and

Adicional al trabajo que ha venido haciendo la Unión Europea (expuesto en otro apartado), la organización que más ha contribuido a la preservación de la privacidad en el ambiente digital y su utilización en sistemas de IA es la *Global Privacy Assembly (GPA)*. La asamblea fue creada en 1979 bajo el nombre de la *International Conference of Data Protection and Privacy Commissioners* para las autoridades de protección de datos y privacidad. En 2003, en Sydney, se adoptaron las primeras cinco resoluciones sobre una combinación de cuestiones que abarcan las transferencias de datos transfronterizas, las cuestiones tecnológicas, la comunicación pública de mensajes de privacidad y las cuestiones de protección de datos dentro de las organizaciones internacionales. La asamblea también busca poner a la privacidad como derecho universal en un mundo globalizado.²⁸¹

En la XLII sesión de la Asamblea Global de Privacidad (octubre de 2020), se emitieron varias resoluciones:

- Resolución sobre tecnología de reconocimiento facial;
- Resolución sobre el papel de la protección de datos personales en la ayuda internacional al desarrollo, la ayuda humanitaria internacional y la gestión de crisis;
- Resolución sobre la rendición de cuentas en el desarrollo y uso de la inteligencia artificial;
- Resolución sobre los desafíos de privacidad y protección de datos derivados de la pandemia COVID-19;
- Resolución sobre declaraciones conjuntas sobre cuestiones emergentes.²⁸²

of the Council on European data governance (Data Governance Act) Version 1.1., junio 2021. https://edpb.europa.eu/system/files/2021-03/edpb-edps_joint_opinion_dga_en.pdf Para conocer, adicionalmente, más sobre el Reglamento de Privacidad Electrónica puede consultarse el siguiente artículo: Reglamento ePrivacy: ¿qué hay que saber?. Digital Guide. IONOS. Marzo 3, 2021. <https://www.ionos.mx/digitalguide/paginas-web/derecho-digital/eprivacy-reglamento-sobre-privacidad-electronica-en-la-eu/> El 21 de abril de 2021 se dio a conocer la “Propuesta de Reglamento sobre un enfoque europeo de la Inteligencia Artificial”, primer marco jurídico sobre IA, que aborda “los riesgos de la IA” en su utilización, consúltese: [Proposal for a Regulation on a European approach for Artificial Intelligence | Shaping Europe’s digital future \(europa.eu\)](https://european-council.europa.eu/media/1234567890/attachment/data/2021/04/Proposal_for_a_Regulation_on_a_European_approach_for_Artificial_Intelligence_-_Shaping_Europe_s_digital_future.pdf)

²⁸¹ *Idem*. El Instituto Nacional de Transparencia y Accesos la Información y Protección de Datos Personales de México participa como miembro de la Asamblea Global de Privacidad.

²⁸² Para conocer el texto de cada una de las resoluciones consúltese: <https://globalprivacyassembly.org/document-archive/adopted-resolutions/>

La “Resolución sobre la rendición de cuentas en el desarrollo y uso de la inteligencia artificial” busca promover la “fiabilidad” de las empresas que desarrollan y utilizan sistemas de IA, para que colaboren estrechamente con los responsables de la formulación de políticas públicas de los gobiernos, para resolver las preocupaciones y corregir los sesgos en materia de Derechos Humanos. Las principales recomendaciones son las siguientes:

- (1) Evaluar el impacto potencial en los Derechos Humanos (incluidos los derechos de protección de datos y privacidad) antes del desarrollo y/o uso de la IA;
- (2) Pruebe la robustez, fiabilidad, precisión y seguridad de los datos de la IA antes de ponerla en uso, incluida la identificación y el abordamiento del sesgo en los sistemas y los datos que utilizan que pueden conducir a resultados injustos;
- (3) Mantener registros de evaluación de impacto, diseño, desarrollo, pruebas y uso de IA;
- (4) Divulgar los resultados de la evaluación de impacto en la protección de datos, la privacidad y los Derechos Humanos de la IA;
- (5) Garantizar la transparencia y la apertura mediante la divulgación del uso de la IA, los datos que se utilizan y la lógica implicada en la IA;
- (6) Asegurar que se identifique a un actor humano responsable (a) con el que se puedan plantear preocupaciones relacionadas con decisiones automatizadas y se puedan ejercer derechos, y b) quién puede desencadenar una evaluación del proceso de decisión y la intervención humana;
- (7) Proporcionar explicaciones en un lenguaje claro y comprensible para las decisiones automatizadas tomadas por la IA a petición;
- (8) Hacer la intervención humana sobre la decisión automatizada tomada por la IA a petición;
- (9) Monitorear y evaluar continuamente el desempeño y los impactos de la IA por parte de los seres humanos, y actuar con prontitud y firmeza para abordar los problemas identificados;
- (10) Implementar mecanismos de denuncia/presentación de informes sobre el incumplimiento o el riesgo significativo en el uso de la IA;
- (11) Garantizar la auditabilidad de los sistemas de IA y estar preparado para demostrar la rendición de cuentas ante las autoridades de protección de datos a petición; Y
- (12) Participar en debates con múltiples partes interesadas (incluidas las organizaciones no gubernamentales, las autoridades públicas y el mundo académico) para identificar y abordar el impacto socioeconómico más amplio de la IA y garantizar la vigilancia algorítmica.
- (13) Instar a las organizaciones que desarrollan o utilizan sistemas de IA a implementar medidas de rendición de cuentas que sean apropiadas con respecto a los riesgos de interferencia en los Derechos Humanos.
- (14) Exhortar a todos los miembros de la Asamblea Mundial de la Privacidad a que colaboren con organizaciones que desarrollen o utilicen sistemas de IA en sus jurisdicciones y a nivel mundial para promover los principios adoptados en su resolución de 2018, y la rendición de cuentas en el desarrollo y uso de la IA, y la adopción de medidas de rendición de cuentas;
- (15) Alentar a los gobiernos a considerar la necesidad de realizar cambios legislativos en las leyes de protección de datos personales, a fin de dejar claras las obligaciones legales relativas a la rendición de cuentas en el desarrollo y uso de la IA, cuando dichas disposiciones aún no estén en vigor; Y
- (16) Alentar a los gobiernos, a los organismos que supervisan o regulan, a las organizaciones que desarrollan o utilizan sistemas de IA y todas las demás partes interesadas pertinentes a que colaboren con las autoridades de protección de datos en el

establecimiento de principios, normas y mecanismos de rendición de cuentas, como la certificación, con el fin de demostrar el cumplimiento legal, la rendición de cuentas y la Ética en el desarrollo y uso de los sistemas de IA.²⁸³

De todo lo anterior, expertos como Carissa Véliz²⁸⁴, han entendido que la regulación es necesaria para hacer respetar el derecho a la privacidad, la libertad, la seguridad jurídica y los Derechos Humanos. Si bien se estuvo dando un fenómeno de la “sociedad de la vigilancia” y de la “economía de la vigilancia”, esto se va a detener con la posibilidad de defender nuestra libertad digital al momento de elegir qué comparto y qué no deseo compartir. Además, la nueva legislación también traerá una nueva ciudadanía que seguramente luchará por oponerse a revelar su información personal y defender sus convicciones por el simple hecho de ser parte de su libertad.

Véliz se pregunta al final de su libro en qué tipo de sociedad nos gustaría vivir. Una sociedad de vigilancia extrema y permanente, en la que no tengamos privacidad ni de parte de los gobiernos ni de parte de las empresas; en el que la vigilancia también se utilice para conocer los gustos de las personas y sus sentimientos, que nos vigila hasta nuestras tendencias políticas y le permite a los más poderosos hacer predicciones para tomar decisiones, en la que las máquinas van administrando la información y te dicen que medites cuando tus niveles de estrés aumentan para que no te conflictúes con la gente. O el segundo tipo de sociedad: “Uno en el que los datos de su teléfono inteligente permanecen allí y nadie tiene acceso a ellos excepto tú.”²⁸⁵

La forma en la que Véliz trata el tema nos recuerda a los filósofos Jeremy Bentham e Isahía Berlin. Del primero, resulta imposible no citar su texto Panóptico cuyo nombre completo, según la portada del libro es *Panopticon; or, the inspection-house: containing the idea of a new principle of construction applicable to any sort of establishment, in which persons of any description are to be kept under inspection;*

²⁸³ Cfr. XLII sesión de la Asamblea Global de Privacidad (octubre de 2020) <https://globalprivacyassembly.org/document-archive/adopted-resolutions/>

²⁸⁴ “Privacy is Power”, Transworld, 2020, Kindel [la traducción es propia]

²⁸⁵ *Idem*.

and in particular to penitentiary-houses, lazarettos, hospitals and schools: with a plan of management adapted to the principle: in a series of letters, written in the year 1787, from crecheff in White Russia, to a friend in England (1791).

El panóptico moderno no tiene un diseño arquitectónico físico, sino un diseño tecnológico que se alimenta con el *Big Data* funciona con la Inteligencia Artificial, se mantiene vivo con el cómputo en la nube y evolucionará con el Internet de las Cosas y la banda 5G. No lo opera un Estado, ni busca asegurar la buena conducta de las personas, su salud, aseo, orden e industria; tampoco pretende prevenir el delito, sino *construir una economía de la vigilancia* a través de los datos que cada uno de nosotros aporta (en la mayoría de las veces sin saber ni entender absolutamente nada sobre su uso, destino o fines).

El panóptico moderno se mantiene funcionando con la regla “ver, sin ser visto” que se producía -según Bentham- debido a la centralidad del inspector; éste sabía todo de los internos, pero los internos desconocían lo que hacía el inspector o su ubicación física. Era un misterio su actuación y era imposible predecir sus acciones, mientras que todo lo que hacían los internos se encontraba bajo la lupa del inspector. Lo mismo ocurre hoy con todas las compañías que utilizan nuestros datos. Mientras no exista un ente público o privado que vigile al inspector, un guardián superior, que bajo el mismo principio de “ver sin ser visto” sea controlado, se generarán abusos y excesos. Eso es natural en cualquier actividad humana o institución que no tiene límites ni es auditada o supervisada.

Véliz lo sabe perfectamente por su formación filosófica y su comprensión del ser humano y el control social y nos muestra las aportaciones de otros filósofos como Michel Foucault, que se mantienen con vigencia y se convierte en parte medular de sus argumentos en diversas partes del libro y al momento del cierre con la frase: “Refuse the unacceptable. Take back control of your personal data, and privacy will prevail.”

El panóptico benthamiano se suma a Isaiah Berlin (1909-1997). Durante las conferencias *Freedom and its Betrayal* (1952) Berlin comentaba: “La esencia de la libertad ha radicado siempre en la posibilidad de elegir lo que se desea elegir,

porque así se desee, sin coerción, sin presiones, sin verse engullido por un vasto sistema; y en el derecho a oponerse, a ser impopular, a defender las convicciones propias simplemente porque son tus convicciones. Ésa es la verdadera libertad y sin ella no hay libertad de ningún tipo, ni siquiera la ilusión de ella.”²⁸⁶

Carissa Véliz encuentra en su libro *Privacidad es Poder* la posibilidad de defender nuestra libertad digital y nuestra privacidad colectiva al momento de elegir qué comparto y qué no deseo compartir. La libertad digital es un concepto multívoco que no se ha podido definir con claridad pero que todos la expresamos con la posibilidad de dar un clic en nuestros dispositivos para decir el monosílabo con el cual Octavio Paz definió a la libertad en el siglo XX: “sí” o un “no” (acepto/no acepto).²⁸⁷ La libertad y, sobre todo, el consentimiento en el mundo digital va a evolucionar ya no como la libertad de los antiguos ni la libertad de los modernos, sino como un nuevo tipo de derecho que estamos tratando de definir, en el que el *ik et nuc* ya no funciona para encontrar el responsable del acto que afecta a una persona.

Horacio Granero, hace una propuesta interesante al decir que existe el orden público tecnológico y el orden público No tecnológico [ya no el tradicional concepto de orden público que conocimos los estudiantes de leyes]. El orden público tecnológico es el conjunto de medidas tomadas por los poderes públicos, tendientes a organizar las relaciones de los usuarios de los bienes y servicios tecnológicos con los responsables de sus desarrollo o producción a fin de evitar un aprovechamiento indebido por parte de éstos en razón de las dificultades generadas por la superioridad de conocimientos y el acceso a la tecnología.”²⁸⁸ En cambio, el orden público No tecnológico es “el conjunto de condiciones fundamentales de vida social

²⁸⁶ Isaiah Berlin, *Freedom and Its Betrayal: Six Enemies of Human Liberty - Updated Edition*, (Hardy H., Ed.). Princeton University Press, Oxford, 1952.

²⁸⁷ Octavio, Paz, “El siglo XX: La experiencia de la libertad, *Vuelta*, número 167, 8 de octubre de 1990.

²⁸⁸ Horacio Granero: *¿Que límites éticos se plantean hoy un hacedor de programas inteligentes?* en el Ciclo de Conferencias sobre Sistemas Inteligentes y Ética, Academia Nacional de la Ciudad de Buenos Aires, 12 de agosto de 2020 https://www.youtube.com/watch?v=hq7_yMp3PrA. Para conocer más detalles sobre este concepto también puede consultarse el libro de Horacio Granero: *El orden público tecnológico*, Educa, Buenos Aires, 2003.

instituidas en una comunidad jurídica, las cuales, por afectar centralmente a la organización de ésta, no pueden ser alteradas por la voluntad de los individuos” ²⁸⁹. En palabras de Granero, el orden público tecnológico “establece en la esfera de la autonomía de la voluntad privada un régimen de limitaciones relativas al contenido de ciertas relaciones jurídicas por razones de bien común.” ²⁹⁰

Se requiere un proceso educativo amplio para llegar a ser una persona que haga lo que Éticamente debe hacerse, en forma voluntaria, habitual, consciente. Esto comienza en la familia, en la escuela y continúa a lo largo de la vida con lo que se conoce como “buenas prácticas” laborales, académicas, científicas, etcétera. Sin embargo, en la medicina, las máquinas y los algoritmos aprenden, almacenan datos, llegan a conclusiones y decide o predicen según los datos que tienen, cuantos más datos se tienen, mayor responsabilidad hay y mayores son las posibilidades para encontrar algún sesgo que vulnere al usuario de esta tecnología, cual, debe ser o han sido -ya en varios casos- sometido a revisión. Esto, traducido a sesgos o errores en el diseño de sistemas inteligentes de apoyo médico puede hacer responsable a varios actores: el creador del sistema, el empresario o fabricante, el vendedor y el cliente o usuario.

Para concluir esta parte de la investigación, expondremos otros temas contenidos en legislación nacional y supranacional, que a julio de 2021 se encontraban en proceso de aprobación o implementación y que buscan regular no solo la transparencia y la privacidad de forma más detallada o específica para el uso de la IA basada en Machine Learning – Deep Learning, sino también otros temas como: 1) rastreo del error y medición del daño, 2) auditabilidad algorítmica, y, 3) medición de riesgos para la protección de los Derechos Humanos.

²⁸⁹ *Idem.*

²⁹⁰ *Idem.*

B) Auditabilidad algorítmica

Para ubicar el tema de la IA basada en Machine Learning – Deep Learning y de la auditabilidad algorítmica es útil recordar que existen tres posibilidades reglamentarias de la misma: 1) Nula legislación jurídica, solo principios éticos y prácticas responsables todo meramente voluntario; 2) Moderada legislación jurídica, que fomente o exija ajustes técnicos que no entren en conflicto significativamente con los beneficios de la IA; o 3) Rigurosa legislación jurídica, que frene o prohíba el despliegue de la tecnología por una imposición de un gobierno.²⁹¹

Las empresas de tecnología, a lo largo del tiempo, han preferido apoyar los dos primeros supuestos y se oponían al tercero ya que consideraban que esa acción afectaría sus ingresos económicos y su crecimiento. El discurso de una IA Ética divulgado por diversas empresas como IBM, Google, Amazon reflejaba precisamente esta posición hasta antes de 2021, fecha en la que públicamente se manifestaron por esperar a que el congreso norteamericano regulase el uso de sistemas de reconocimiento facial en labores de seguridad pública.

El diseño basado en la Ética ha llevado también a los científicos informáticos a desarrollar investigaciones propias sobre “equidad algorítmica” (concepto acuñado en el año 2012 por investigadores de Microsoft, IBM, y la Universidad de Toronto) y auditorías algorítmicas con un enfoque basado en riesgos (concepto traído del sector financiero). También existen propuestas de auditorías para medir impactos a Derechos Humanos, cuya amplitud y profundidad van más allá del Enfoque Basado en Riesgos que se utiliza por oficiales de cumplimiento del sector financiero y de las llamadas actividades vulnerables. Tómese como ejemplo de este tipo de auditorías la propuesta que hacen A. Mantelero y M. S. Esposito: “*An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems*”(Computer Law & Security Review, 2021) en la que deja a un lado el enfoque de los datos personales (privacidad y

²⁹¹ Dwork, Cynthia, Hardt, Moritz, et. Al. Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12). Association for Computing Machinery, New York, NY, USA, 214–226. DOI: <https://doi.org/10.1145/2090236.2090255>

transparencia) y el enfoque ético (opacidad, interpretabilidad) que explicamos en nuestros apartados previos para abordar los desafíos de la IA basada en Machine Learning – Deep Learning bajo una tercera vía que toma como eje el papel que juegan los Derechos Humanos en la vida de las personas y el impacto que tiene la regulación de los sistemas inteligentes. Lo relevante del documento es que muestra, a través del análisis de más de 700 decisiones y documentos de las autoridades de protección de datos de seis países, como los Derechos Humanos ya sustentan las decisiones de estas autoridades y, por razones obvias, serán los que se utilicen ante algún conflicto judicial que exija la reparación de daño ya no por un hecho ilícito civil o penal, sino por afectación a Derechos Humanos, que de suyo puede generar violaciones directas al régimen constitucional de un Estado.

Universidades como el MIT y Harvard, para tratar el tema desde un enfoque ético y legal, han recibido patrocinios de la industria tecnológica para crear institutos especializados en IA. Por ejemplo, el Data & Society Research Institute (<https://datasociety.net>) está dirigido por un investigador de Microsoft y financiado por una subvención de Microsoft; El Instituto AI Now (<https://ainowinstitute.org>) de la Universidad de Nueva York fue cofundado por otro investigador de Microsoft y financiado parcialmente por Microsoft, Google y *DeepMind*; el Stanford Institute for Human-Centered AI está codirigido por un ex vicepresidente de Google; la División de Ciencias de Datos de la Universidad de California en Berkeley está dirigida por un ex empleado de Microsoft; y el MIT Schwarzman College of Computing está encabezado por un miembro de la junta directiva de Amazon.²⁹²

¿Cuál es el estado del arte en la creación de legislación? Existen varios países que están realizando diversas acciones para entender los beneficios de la IA y los riesgos que se deben regular. De la revisión hecha resalta la postura de Chile, quien ha decidido regular la IA a nivel constitucional y se encuentra en puerta la aprobación de la llamada reforma de neuroderechos (y una ley secundaria para su desarrollo). Los neuroderechos que serán plasmados en la constitución chilena

²⁹² Cfr. <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>

siguen las ideas e investigaciones del doctor Rafael Yuste y se concentran en 5 tipos:

1. Derecho a la privacidad mental (los datos cerebrales de las personas);
2. Derecho a la identidad y autonomía personal;
3. Derecho al libre albedrío y a la autodeterminación;
4. Derecho al acceso equitativo a la aumentación cognitiva (para evitar producir inequidades); y,
5. Derecho a la protección de sesgos de algoritmos o procesos automatizados de toma de decisiones.

Desde 2017, Yuste se ha manifestado en defender la identidad y la autonomía personal como prioridades Éticas de la neurociencia y la IA, recomendando incluir en las legislaciones del mundo cláusulas que protejan los neuroderechos y sostiene que se debería crear una convención internacional para que los seres humanos conozcan sus derechos y se ponga límites a su uso en casos que afecten a la persona, su entorno o a la sociedad en conjunto. Yuste conjuntamente con un equipo de científicos, comentan en la revista Nature:

Un primer paso hacia [la regulación] sería exponer a los ingenieros, otros desarrolladores de tecnología y aprendices de investigación académica a ***la Ética como parte de su capacitación estándar para unirse a una empresa o laboratorio***. Se podría enseñar a los empleados a pensar más profundamente sobre cómo perseguir avances y desplegar estrategias que puedan contribuir de manera constructiva a la sociedad, en lugar de fracturarla. ***Este tipo de enfoque seguiría esencialmente al utilizado en medicina***. A los estudiantes de medicina se les enseña sobre la confidencialidad del paciente, la no lesión y sus deberes de beneficencia y justicia, y deben tomar el Juramento Hipocrático para adherirse a los más altos estándares de la profesión. Los posibles beneficios clínicos y sociales de las neurotecnologías son enormes. Para cosecharlos, debemos orientar su desarrollo de una manera que respete, proteja y habilite lo mejor de la humanidad.²⁹³

²⁹³ Cfr. Yuste, R., Goering, S., Arcas, B. et. Al., “Four ethical priorities for neurotechnologies and AI” en *Nature*, 551, 159–163, 2017. Consúltase también el texto en: <https://doi.org/10.1038/551159a> <https://www.nature.com/articles/551159a.pdf>

Existe otro ejemplo en Francia, donde encontramos el “PEReN” o “Pôle d’Expertise et de Régulation Numérique” es una unidad administrativa instalada en Ivry-sur-Seine (Val-de-Marne) para analizar, entender, supervisar y controlar a estos grandes colosos tecnológicos. Fue creado por los ministerios de economía, tecnología digital y cultura para ayudar al Estado, pero también la Autoridad de Competencia, el Consejo Superior Audiovisual (CSA), el regulador de telecomunicaciones (ARCEP) y la Comisión Nacional de Informática y Libertades (CNIL).

Según Alexandre Piquard, la principal justificación de su creación la encontramos en algunas preguntas que están por resolverse con esta creación: ¿cómo regular las plataformas digitales sin poder evaluar su ranking o algoritmos de recomendación? ¿Cómo comprobar que estos actores respetan las leyes? ¿Cómo garantizar que un motor de búsqueda, una tienda de aplicaciones o un mercado no abusen de su posición dominante sin auditar los resultados de sus páginas? Thomas Courbe, director en la Dirección General de Empresas, del servicio de Bercy que alberga el PEReN explica: “Hay una apuesta muy fuerte, porque varias normas prevén dar a las autoridades una capacidad de auditoría de algoritmos”, por ejemplo: la Ley de Mercados Digitales y la Ley de Servicios Digitales, que en Francia ya tiene avances en diversos proyectos de ley para que se respeten los principios del estado francés.²⁹⁴

²⁹⁴ “Le PEReN a été créé par les ministères de l’économie, du numérique et de la culture pour aider l’Etat, mais aussi l’Autorité de la concurrence, le Conseil supérieur de l’audiovisuel (CSA), le régulateur des télécoms (Arcep) ou la Commission nationale de l’informatique et des libertés (CNIL). Tous sont confrontés à des questions brûlantes : comment réguler les plates-formes numériques sans pouvoir évaluer leurs algorithmes de classement ou de recommandation ? Comment vérifier que ces acteurs respectent les lois? Comment s’assurer qu’un moteur de recherche, un magasin d’applications ou une place de marché n’abuse pas de sa position dominante sans auditer les résultats sur ses pages?” Cfr. Alexandre Piquard, A Bercy, une cellule d’informaticiens pour aider l’Etat à réguler les GAFA” *LeMonde*, 22 de enero de 2021 en https://www.lemonde.fr/economie/article/2021/01/22/a-bercy-une-cellule-d-informaticiens-pour-aider-l-etat-a-reguler-les-gafa_6067191_3234.html Francia prohibió analizar los datos de los tribunales y predecir a los jueces para proteger la independencia judicial: Article 33 - LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice (1) - Légifrance (legifrance.gouv.fr)

Sin embargo, para efectos de esta investigación, se consideró prudente limitar el estudio a la legislación de los Estados Unidos (por ser mayor productor de IA, el que tiene a las empresas más importantes que generan IA basada en Machine Learning – Deep Learning) y a la Unión Europea (por ser el principal consumidor de IA en el mundo).²⁹⁵

En los Estados Unidos. En el Congreso de los Estados Unidos existen varias iniciativas relacionadas con el uso de algoritmos y auditabilidad algorítmica. En particular mencionaremos dos que consideramos de relevancia:

1. Iniciativa de Ley de Responsabilidad Algorítmica de 2019²⁹⁶: Busca que las empresas que utilizan sistemas de inteligencia artificial, cuando involucren información personal o tomen decisiones automatizadas, realicen evaluaciones de sus algoritmos. Específicamente, los sistemas de decisión automatizados “de alto riesgo” como por ejemplo aquellos que pueden contribuir a la inexactitud, el sesgo o la discriminación; o aquellos que pueden facilitar la toma de decisiones sobre aspectos sensibles de la vida de las personas mediante la evaluación de su comportamiento. Un sistema de decisión automatizada, o un sistema de información que involucre datos personales, se considera de alto riesgo si: 1. plantea problemas de seguridad o privacidad; 2. involucra la información personal de un número significativo de personas; o 3. sistemáticamente supervisa una gran ubicación física de acceso público. Las evaluaciones de los sistemas de decisión automatizada de alto riesgo deben describir el sistema a detalle, evaluar los costos y beneficios relativos del sistema, determinar los riesgos para la privacidad y seguridad de la información personal, y explicar las medidas adoptadas para minimizar esos

²⁹⁵ Aunque países como Chile que está tomando muy en serio estos temas, al grado estar a punto de reformar su Constitución para incluir los llamados neuroderechos (1. Derecho a la privacidad mental (los datos cerebrales de las personas); 2. Derecho a la identidad y autonomía persona; 3. Derecho al libre albedrío y a la autodeterminación; 4. Derecho al acceso equitativo a la aumentación cognitiva (para evitar producir inequidades); y, el que nos interesa a nosotros -una vez que exista legislación secundaria- *Derecho a la protección de sesgos de algoritmos o procesos automatizados de toma de decisiones*) Y Francia con Tomaremos para efectos de este trabajo se consideró de relevancia limitar el estudio a la legislación de los Estados Unidos (por ser un país productor de IA, el mayor productor del mundo) y la legislación de la Unión Europea (por ser el principal consumidor de IA en el mundo).

²⁹⁶ Consúltese la iniciativa en: <https://www.govtrack.us/congress/bills/116/s1108/text>

riesgos, si es que se llegan a descubrir. Las evaluaciones de los sistemas de información de alto riesgo que involucran información personal deben evaluar hasta qué punto el sistema protege la privacidad y seguridad de dicha información.

2. Iniciativa de Ley de Equidad Algorítmica de 2020²⁹⁷: Busca cuidar e incrementar la equidad y transparencia en las decisiones algorítmicas de elegibilidad. Incorpora conceptos como “equidad sustantiva” “equidad procesal” y “determinación algorítmica de elegibilidad”. Esta última, definida como una determinación basada en un proceso algorítmico que utiliza métodos de aprendizaje automático, técnicas estadísticas avanzadas, inteligencia artificial o técnicas similares para determinar la elegibilidad, la negativa de acceso al costo de adquisición, oportunidad de acceso o revocación de oportunidades importantes que incluyen, entre otras, educación, empleo, crédito, seguro médico y vivienda.

A nivel local, en el Ayuntamiento de Nueva York, se han presentado cuatro proyectos de ley para regular los algoritmos y las decisiones de sistemas automatizados, también se está buscando regular la venta de herramientas automatizadas de decisión para empleos. Este último proyecto de ley prohíbe la venta/compra de tales herramientas sino se demuestra que fueron objeto de una “auditoría por sesgo” durante el último año antes de la venta/compra²⁹⁸. La siguiente tabla contiene las cuatro propuestas de ley presentadas en el ayuntamiento de Nueva York:

²⁹⁷ Consúltase en: <https://www.govtrack.us/congress/bills/116/s5052/text/is>

²⁹⁸ Cfr. [The New York City Council - File #: Int 1894-2020 \(nyc.gov\)](#) Este proyecto de ley también requeriría que cualquier persona que utilice herramientas automatizadas de evaluación del empleo para la contratación y otros fines de empleo se divulgue a los candidatos, en un plazo de 30 días, cuando dichas herramientas se utilizaron para evaluar su candidatura para el empleo, y las cualificaciones o características laborales para las que la herramienta se utilizó para examinar.

Proyectos legislativos en el ayuntamiento de Nueva York que regulan los algoritmos o la IA

No. de Archivo	No. Ley	Estado	Promovente	Tema
Int 1894-2020		En comisión	Laurie A. Cumbo	Modificar el código administrativo de la ciudad de Nueva York, en relación con la venta de herramientas automatizadas de decisión para el empleo
Int 1806-2019		En comisión	Peter A. Koo	Modificar el código administrativo de la ciudad de Nueva York, en relación con la presentación de informes sobre los sistemas de decisión automatizados utilizados por las agencias de la ciudad
Int 1696-2017	2018/049	Aprobada	James Vacca	Regular los sistemas automatizados de decisión utilizados por las agencias
Int 0029-2010	2012/011	Aprobada	Gale A. Brewer	Una Ley Local para modificar el código administrativo de la ciudad de Nueva York, en relación con la publicación de datos abiertos.

Fuente: Elaboración propia con base en información obtenida en: [The New York City Council – Legislation \(nyc.gov\)](https://www.nyc.gov/legislation)

Por otro lado, la Comisión Federal de Comercio (FCT) está investigando sobre las prácticas de privacidad de algunas de las empresas de tecnología y plataformas de redes sociales más grandes del mundo (Amazon, la empresa matriz de TikTok, ByteDance, Discord, Facebook, Reddit, Snap, Twitter, WhatsApp y YouTube), para que proporcionen datos sobre cómo recopilan y utilizan información personal de sus usuarios.²⁹⁹ Si bien no habla de algoritmos como tal, los gigantes comerciales van a tener que construir áreas especializadas para estos temas con equipos multidisciplinarios que permitan atender las exigencias de las autoridades y los ciudadanos. Por su parte, los gobiernos, deberán valorar la creación de unidades administrativas especializadas en estos temas. En conjunto, se deberá trabajar para cuidar al ciudadano. Si uno no lo hace y no entiende esta necesidad, habrá fallas en la protección de Derechos Humanos.

²⁹⁹ Makena Kelly, "The FTC is investigating data collection at YouTube, Facebook, and seven other companies", *The Verge*, 14 diciembre 2020, <https://www.theverge.com/2020/12/14/22174755/ftc-data-privacy-youtube-facebook-reddit-twitter-snapchat-discord-inquiry>

Durante 2018, ocurrieron varios acontecimientos relevantes que son producto del uso de los algoritmos. Por ejemplo: Microsoft publicó sus principios éticos para la IA, haciendo énfasis en la equidad (enero). Facebook anunció su compromiso con el desarrollo ético y el despliegue de la IA y una herramienta para buscar sesgo llamada Flujo de Equidad (mayo). Google publicó sus “prácticas responsables” para la investigación y el desarrollo de la IA (junio). IBM anunció una herramienta llamada AI Fairness 360, diseñada para comprobar si hay sesgos no deseados en datasets y modelos de aprendizaje automático (septiembre).³⁰⁰

En 2018, *Reuters* informó que *Amazon* eliminó una herramienta que filtraba currículums basados en patrones de contratación que discriminaba a las mujeres³⁰¹. Las empresas y organizaciones de la sociedad civil han están enfocando la discusión hacia el tema de principios éticos y el uso de la IA mediante prácticas responsables ubicando términos como sesgo y equidad.

En la Unión Europea. En este espacio encontramos varios ejemplos de normas que buscan la regulación algorítmica: 1) “Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial (2020/2014(INL)), P9_TA(2020)0276; y, 2) Resolución del Parlamento europeo, de 20 de enero de 2021, sobre inteligencia artificial: cuestiones de interpretación y de aplicación del Derecho Internacional en la medida en que la UE se ve afectada en los ámbitos de los usos civil y militar, así como de la autoridad del Estado fuera del ámbito de la justicia penal (2020/2013(INI)), P9_TA-PROV(2021)0009.

Adicionalmente, el 21 de abril de 2021, en el Parlamento Europeo, fue presentada la “*Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and*

³⁰⁰ Cfr. Rodrigo, Ochigame, “THE INVENTION OF “ETHICAL AI. How Big Tech Manipulates Academia to Avoid Regulation” en *The Intercept*, 20 de diciembre, 2019. <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence> [15/02/2021]

³⁰¹ Jeffrey Dastin, “Amazon scraps secret AI recruiting tool that showed bias against women”, *Reuters*, 10 de octubre de 2018, en <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

Amending Certain Union Legislative Acts".³⁰² En su exposición de motivos se deja claro que si bien la mayoría de los sistemas de IA plantearán un riesgo bajo o nulo, algunos de esos sistemas entrañan peligros que deben regularse por los gobiernos para evitar situaciones no deseadas. Existen ya casos que han sido detectados que se pueden clasificar como nuevos riesgos relacionados con la seguridad de los usuarios y los derechos fundamentales. Esto afecta la seguridad jurídica y la posibilidad de que las personas y empresas adopten la IA de manera pronta debido a la falta de confianza, proponiendo cuatro niveles de clasificación basada en riesgos. Por su relevancia se transcribe a continuación:

"La Comisión propone un **planteamiento basado en el riesgo**, con cuatro niveles de riesgo: **Riesgo inadmisibile**: Se prohibirá un conjunto muy limitado de usos especialmente nocivos de la IA que contravienen los valores de la Unión al violar los derechos fundamentales (por ejemplo, puntuación social por parte de los Gobiernos, explotación de los puntos débiles de los niños, uso de técnicas subliminales y, salvo contadas excepciones, determinados sistemas de identificación biométrica remota en directo en espacios públicos con fines policiales). **Alto riesgo**: Se considera de alto riesgo un número limitado de sistemas de IA definidos en la propuesta y que tienen un impacto negativo en la seguridad de las personas o en sus derechos fundamentales (protegidos por la Carta de los Derechos Fundamentales de la UE). Se adjunta a la propuesta la lista de sistemas de IA de alto riesgo, que puede revisarse para adaptarla a la evolución de los casos de uso de la IA (a prueba de futuro). Entre ellos figuran los componentes de seguridad de los productos contemplados en la legislación sectorial de la Unión. Siempre serán de alto riesgo cuando estén sujetos a una evaluación de la conformidad por terceros con arreglo a dicha legislación sectorial. A fin de garantizar la confianza y un nivel elevado y coherente de protección de la seguridad y los derechos fundamentales, se proponen requisitos obligatorios para todos los sistemas de IA de alto riesgo. Esos requisitos se refieren a la calidad de los conjuntos de datos utilizados; la documentación técnica y la llevanza de registro; la transparencia y la divulgación de información a los usuarios; la supervisión humana, y la solidez, la precisión y la ciberseguridad. En caso de infracción, los requisitos permitirán a las autoridades nacionales acceder a la información necesaria para investigar si el uso de la IA cumplió la legislación. La normativa propuesta es coherente con la Carta de los Derechos Fundamentales de la Unión Europea y está en consonancia con los compromisos internacionales de la Unión en materia de comercio. **Riesgo limitado**: En el caso de determinados sistemas de IA se imponen obligaciones específicas de transparencia, por ejemplo, cuando exista un riesgo claro de manipulación (por ejemplo, mediante el uso de robots conversacionales). Los usuarios deben ser conscientes de que están interactuando con una máquina. **Riesgo mínimo**: Todos los demás sistemas de IA pueden desarrollarse y utilizarse con arreglo a la legislación vigente sin obligaciones jurídicas adicionales. La inmensa mayoría de los sistemas de IA utilizados actualmente en la UE pertenece a esta categoría. De forma voluntaria, los proveedores de estos sistemas pueden optar por aplicar los requisitos de una IA digna de confianza y adherirse a códigos de conducta voluntarios."³⁰³

³⁰² Consúltese el texto en: https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683

³⁰³ Consúltese el texto en: https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683

Para los efectos de nuestra investigación nos interesan los sistemas de alto riesgo ya que ahí se encuentra la IA basada en Machine Learning – Deep Learning enfocada a la salud. Una IA de riesgo alto incluyen la tecnología de IA utilizada en:

- Infraestructuras críticas (por ejemplo, transporte), que podrían poner en riesgo la vida y la salud de los ciudadanos;
- Formación educativa o vocacional, que puede determinar el acceso a la educación y el curso profesional de la vida de alguien (por ejemplo, calificación de exámenes);
- Componentes de seguridad de los productos (por ejemplo, aplicación de IA en cirugía asistida por robot);
- Empleo, gestión de trabajadores y acceso al trabajo por cuenta propia (por ejemplo, software de clasificación de CV para los procedimientos de contratación);
- Servicios públicos y privados esenciales (por ejemplo, calificación crediticia que niega a los ciudadanos la oportunidad de obtener un préstamo);
- Aplicación de la ley que pueda interferir con los derechos fundamentales de las personas (por ejemplo, evaluación de la confiabilidad de las pruebas);
- Gestión de migración, asilo y control de fronteras (por ejemplo, verificación de la autenticidad de los documentos de viaje);
- Administración de justicia y procesos democráticos (por ejemplo, aplicación de la ley a un conjunto concreto de hechos).

Los sistemas de IA de alto riesgo estarán sujetos a estrictas obligaciones antes de que puedan comercializarse:

- Sistemas adecuados de evaluación y mitigación de riesgos;
- Alta calidad de los conjuntos de datos que alimentan el sistema para minimizar los riesgos y los resultados discriminatorios;
- Registro de actividad para garantizar la trazabilidad de los resultados;

- Documentación detallada que proporcione toda la información necesaria sobre el sistema y su propósito para que las autoridades evalúen su cumplimiento;
- Información clara y adecuada al usuario;
- Medidas apropiadas de supervisión humana para minimizar el riesgo;
- Alto nivel de robustez, seguridad y precisión.³⁰⁴

Este proyecto de creación de una legislación específica para el uso de la IA marcará la pauta de regulación en gran parte del mundo ya que su cumplimiento debe hacerse por personas físicas y jurídicas (públicas y privadas) de dentro y fuera de la UE en la medida en que el sistema de IA se introduzca en el mercado de la Unión o su uso afecte a personas establecidas en ella (usuarios, proveedores, programadores), siendo su única excepción los usos privados no profesionales. América Latina, China y Estados Unidos, que tienen relaciones comerciales y de negocios con la UE deberán ir analizando la manera de incorporar algunos ajustes a su legislación. La norma da la pauta también para construir un ecosistema de protección social por el uso de la IA que en su momento explicaremos con mayor detalle.³⁰⁵

¿Quién más participa en el análisis y la creación de legislación? En 2017, la Universidad de New York, creó el Instituto *AI Now*³⁰⁶ como un centro de investigación interdisciplinario dedicado a comprender las implicaciones sociales de la inteligencia artificial para garantizar que los sistemas de IA sean sensibles a las complejas necesidades sociales en los que se aplican y desarrollar nuevas formas de medirlos, auditarlos, analizarlos o mejorarlos. En octubre de 2018 publica *Algorithmic Accountability Policy Toolkit* como una herramienta novedosa para proveer a los defensores jurídicos una comprensión básica sobre el uso

³⁰⁴ Consúltase el texto en: https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683

³⁰⁵ La UE tiene la intención de profundizar en las asociaciones, coaliciones y alianzas con socios de la UE (por ejemplo, Japón, los Estados Unidos o la India), así como multilaterales (por ejemplo, OCDE y G-20) y organizaciones regionales (por ejemplo, el Consejo de Europa).

³⁰⁶ [AI Now Institute](#)

gubernamental de los algoritmos, incluido un desglose de los conceptos clave y las preguntas que pueden surgir al abordar este tema, así como una descripción general de las investigaciones existentes y resúmenes de los sistemas algorítmicos.³⁰⁷

En mayo de 2018, Amnistía Internacional, Access Now y otras organizaciones asociadas presentaron la Declaración de Toronto sobre la protección del derecho a la igualdad y la no discriminación en los sistemas de aprendizaje automático. La Declaración es un documento que busca aplicar las normas de Derechos Humanos existentes al uso de sistemas de aprendizaje automático.³⁰⁸

El 28 de febrero de 2020, en el Vaticano, se reunieron el presidente de IBM (John Kelly III), el presidente de Microsoft (Brad Smith), el director de la FAO (Qu Dongyu), el presidente de la Academia de la Vida (Vincenzo Paglia) y la Ministra Italiana para la Tecnología, Innovación y digitalización (Paola Pisano), teniendo como testigo al presidente del Parlamento Europeo (David Sassoli) para firmar el documento “Rome call for AI Ethics”.³⁰⁹ El objetivo de la reunión fue mantener a la IA con un sentido humanista. Es decir, respetar la Declaración Universal de los Derechos Humanos, buscar el desarrollo de todas las personas, cuidar su dignidad y el medio ambiente, así como las necesidades de los más vulnerables.

Los hechos que genere la IA deben ser el punto de referencia para una Ética digital y la promoción del uso de la tecnología en beneficio de a la humanidad y el medio ambiente. Todos los seres humanos nacen libres e iguales en dignidad y derechos. Están dotados de razón y conciencia, deben comportarse fraternalmente los unos con los otros (art.1 DDHH). Esta condición fundamental de libertad y dignidad -dice la declaración- debe ser también protegida y garantizada cuando se producen y usan los sistemas de IA.

³⁰⁷ Cfr. “Algorithmic Accountability Policy Toolkit” *op. cit.*, Consúltese en: <https://ainowinstitute.org/aap-toolkit.pdf>

³⁰⁸ The Toronto Declaration: Protecting the rights to equality and non-discrimination in *Machine Learning* systems: <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/>

³⁰⁹ Rome Call for AI Ethics se puede consultar en: https://www.romecall.org/wp-content/uploads/2021/02/AI-Rome-Call-x-firma_DEF_DEF_con-firme_.pdf

Para ello deben existir tres requisitos: 1) Incluir a todo ser humano, sin ninguna discriminación; 2) debe hacer cosas buenas para la humanidad y para cada ser humano; y 3) debe ser consciente de la realidad compleja de nuestro ecosistema y estar definida por el modo en el cual cuida y protege el planeta (uso sustentable). La obligación de hacer sistemas transparentes/explicables es un criterio básico de la IA y de los algoritmos de agentes inteligentes. Los principios contenidos en esta declaración son:

- *Transparency* (transparencia): en principio, los sistemas de IA deben ser explicables.
- *Inclusion* (inclusión): La necesidad de que todos los seres humanos sean tomados en cuenta para que cada uno pueda beneficiarse y que a todos los individuos le sean ofrecidas las mejores condiciones para expresarse y desarrollarse.
- *Responsibility* (responsabilidad) aquellos que diseñan y despliegan el uso de la IA deben proceder con responsabilidad y transparencia.
- *Impartiality* (imparcialidad) no crear o actuar de acuerdo con sesgos, sino salvaguardar la justicia y la dignidad humana;
- *Reliability* (fiabilidad) los sistemas de IA deben ser capaces de trabajar fiablemente;
- *Security and privacy* (seguridad y privacidad): los sistemas de IA deben trabajar con seguridad y respeto por la privacidad de los usuarios.³¹⁰

En 2020, IEEE Standards Association (IEEE SA), desarrolló un estándar que busca medir el impacto de la inteligencia artificial o los sistemas autónomos e inteligentes (A / IS) en el bienestar de los seres humanos, así como los vínculos y conexiones entre el bienestar individual y colectivo, partiendo de los Derechos

³¹⁰ El texto original dice: "For this reason, the sponsors of the call express their desire to work together, in this context and at anational and international level, to promote "algor-ethics", namely the ethical use of AI as definedby the following principles:1.Transparency: in principle, AI systems must be explainable;2.Inclusion: the needs of all human beings must be taken into consideration so that everyone can benefit and all individuals can be offered the bestpossible conditions to express themselves and develop;3.Responsibility: those who design and deploy the use of AI must proceed with responsibilityand transparency;4.Impartiality:do not create or act according to bias, thus safeguarding fairness and human dignity;5.Reliability: AI systems must be able to work reliably; 6.Security and privacy:AI systems must work securely and respect the privacy of users."

Humanos, la equidad e igualdad económicas, la igualdad social, la salud del medioambiente, la estabilidad laboral de las personas, los gobiernos honestos y confiables, y otras dimensiones conectadas con el bienestar individual y colectivo. La IEEE SA reconoce la importancia del Reglamento General de Protección de Datos (RGPD) como referencia a seguir cuando se utiliza el IEEE Std 7010 y adicionalmente proporciona otros recursos (códigos, pautas procesales, estándares) que se relacionan con el uso de datos pero que están fuera del alcance de IEEE Std 7010. Entre los aspectos que podrían considerarse como negativos del uso de A / IS señala: "... Impactos económicos y laborales, incluido el desplazamiento laboral, el desempleo y la desigualdad; responsabilidad, transparencia y explicabilidad; vigilancia, privacidad y libertades civiles; equidad, Ética y Derechos Humanos; manipulación política, engaños en publicidad; salud física y psicológica humana; impactos ambientales; dignidad humana, autonomía humana versus A / IS; y seguridad, ciberseguridad y armas autónomas"³¹¹

En noviembre de 2021, la UNESCO concluyó la redacción de su recomendación sobre la Ética de la Inteligencia Artificial, en la que participaron todos los Estados miembros, las principales empresas de tecnología (Google, Facebook y Microsoft), y universidades como Stanford y Nueva York; también estuvo China con su Academia de Ciencia y Tecnología. El objeto de la Declaración sobre la Ética de la Inteligencia Artificial es sentar las bases para que los sistemas de inteligencia artificial trabajen por el bien de la humanidad. "Entre las preocupaciones que han conducido a la Unesco a impulsar este proyecto figura el potencial de la inteligencia artificial para exacerbar las desigualdades globales. Hoy, solo doscientas firmas producen el 77% de las nuevas patentes e innovaciones relacionadas con el mundo digital. De las 15 principales, 11 están en los Estados Unidos y el resto están en China. La postura de la entidad es que estas tecnologías emergentes deben desarrollarse de forma más inclusiva y compartida. "Los Estados miembro deben

³¹¹ Cfr. Standards Committee of the IEEE Systems, Man, and Cybernetics Society. (2020). *IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being* (Aprobado 5 March 2020/ Publicado 1 de mayo 2020). IEEE SA Standards Board. doi: 10.1109/IEEESTD.2020.9084219
<https://standards.ieee.org/content/ieee-standards/en/standard/7010-2020.html>

diseñar mecanismos que prevengan la monopolización de sistemas de inteligencia artificial... la computación requerida para la investigación en *Deep Learning* se ha multiplicado por 300.000 entre 2012 y 2018. No en vano, el entrenamiento de un gran modelo de IA puede producir más de 270 toneladas de emisiones de CO₂, cinco veces la cantidad que produce un coche medio durante su vida útil.”³¹²

En palabras de Gabriela Ramos, subdirectora General de Ciencias Sociales y Humanas de la OCDE: “En cuestiones tecnológicas, la Ética lo es todo y debe prevalecer en todo... Cuando estaba trabajando en [los principios de la OCDE] siempre les pregunté: “Oigan, ¿dónde está la Ética?”. Porque ya estaban los que todos conocemos, que no es la primera vez que se habla de ellos: transparencia, rendición de cuentas, responsabilidad, seguridad... Y siempre me dijeron que la Ética la tiene la Unesco. Me pareció una cosa muy loca pensar que la Ética era privativa de una sola institución. La Ética es todo, debe prevalecer en todo y debe informar todo.”³¹³

Todas estas propuestas para la regulación de los algoritmos buscan auditar las consecuencias negativas de su uso. Cabe señalar que la creación de legislación sin generar mecanismos de supervisión es, en muchos temas, letra muerta. Por ello, coincidimos con aquellos autores que están proponiendo la creación de empresas que realicen auditorías algorítmicas de IA basada en Machine Learning – Deep Learning y también de agencias públicas reguladoras de estos temas. Sin lugar a duda este tipo de acciones de supervisión ha traído grandes resultados a muchos sectores de la producción y los servicios. Sobre todo, en el ámbito financiero y de los servicios de salud. Partamos de una pregunta ¿Cómo determinar en una auditoría algorítmica qué es bueno y qué es malo para el ser humano? Como hemos dicho, en esto ha habido varios caminos. El camino de los datos y la privacidad, el camino de la Ética de la IA y, un tercer camino está en la protección de los Derechos Humanos. Para llegar a esto último primero analizaremos las propuestas de

³¹² Cfr. [La Unesco pone los cimientos de la 'Declaración Universal de la inteligencia artificial' | Tecnología | EL PAÍS \(elpais.com\)](#)

³¹³ Cfr. [“En cuestiones tecnológicas, la Ética lo es todo y debe prevalecer en todo” | Tendencias | EL PAÍS Retina \(ampproject.org\)](#)

auditorías algorítmicas basadas en un enfoque de riesgos y concluiremos explicando la metodología de Evaluación de Impacto en Derechos Humanos (EIDH).

Creemos que las declaraciones de Derechos Humanos son el punto de partida para la toma de decisiones. En este sentido, una aplicación Ética implica poner en el centro al ser humano y sus derechos (dignidad, no discriminación, cuidado del medio ambiente; ayuda a los más débiles; transparencia en las decisiones, inclusión, entre otros).

Algunos han manifestado que se debe poner un protocolo de gestión de disputas y un modelo del control y gestión de riesgos para mitigar las amenazas naturales de la tecnología y su uso en el mundo y, a la vez, la exigencia a las empresas que trabajan con IA de un Certificado de Integridad Ética para demostrar que están cumpliendo los estándares que cada país o región determine. Esto obliga a trabajar con modelos de medición como las matrices de riesgo y sobre programas de responsabilidad basada en riesgos (risk-based liability) o Accountability – Due Diligence.

“La implementación de un sistema de gestión de riesgos parametrizado en un *Blockchain* con algoritmo Pol, que registra la actividad de la AI desde la programación y configuración de fábrica y toda la personalización del usuario, permitiría identificar responsabilidades para efectos de acciones civiles reparativas o de consumidor y gestión de seguros.”.³¹⁴ En palabras de Rafael Y. Cuartas Báez: “Una IA autorregulada mediante *smartcontracts* que conforman un sistema de gestión de riesgos programados con estándares éticos y técnicos internacionales reconocidos por el regulador; que permita interoperabilidad, controles preventivos con alertamiento e identificación de responsabilidad ante la materialización de riesgos. Sistema totalmente auditable y transparente para el usuario, el supervisor

³¹⁴ Rafael Y. Cuartas Báez, “La Blockchain como elemento de gobernanza, confianza, Ética y autorregulación de la IA” en XVI Conferencia del Ciclo Sistemas Inteligentes y Ética, Academia nacional de Ciencias de Buenos Aires, 11 de noviembre 2020. Cfr. https://www.youtube.com/watch?v=6Am_j6uXmYc [1:49:15]

e interesados, con registros o logs inalterables que permite entender la operación de la IA y retroalimentación para ajustar la matriz.”³¹⁵

La rendición de cuentas algorítmicas tiene que ver con la posibilidad de que los algoritmos sean auditables en su construcción y en las decisiones que toman o los problemas que solucionan. La auditabilidad algorítmica debe estar presente al momento que se pone el sistema de inteligencia artificial en operación (en línea o en producción) y durante toda su etapa de funcionamiento y actualización; también cuando ocurre una afectación a algún usuario o a alguna tercera persona como herramienta para conocer la responsabilidad de las personas.

La transparencia en los algoritmos es un elemento sin el cual no es posible realizar una rendición de cuentas algorítmicas (auditoría algorítmica). Es decir, sino se conocen los valores, el resultado, los datos y modelo predictivo y la forma en cómo se calibra y se actualiza el algoritmo, no es posible ubicar la responsabilidad algorítmica. La explicabilidad de un algoritmo juega un papel clave en la detección, habilitación y mejora de la auditabilidad, equidad, confianza y confiabilidad.

El análisis del nexo entre transparencia y rendición de cuentas solo puede proceder después de la introducción de tres distinciones importantes: 1) se deben distinguir varias fases de la toma de decisiones: la recopilación de datos (conjunto de datos), posteriormente los datos se utilizan en el aprendizaje automático para desarrollar un modelo (desarrollo del modelo), y finalmente ese modelo se utiliza para la toma de decisiones (uso del modelo). 2) se deben distinguir los posibles destinatarios (o beneficiarios) de la divulgación: organismos intermedios que tienen algún papel de supervisión, individuos afectados o el público en general, que obviamente incluye las dos categorías anteriores. 3) En conjunto, estas dos distinciones implican que cualquier llamado a la transparencia debe ser específico sobre lo que se debe revelar y a quién se debe dar. El aprendizaje automático está muy relacionado entre sí en las etapas de recopilación de datos, construcción del modelo y uso del modelo. Cuando se llama a una organización a rendir cuentas se debe analizar todo.

³¹⁵ *Idem.*

La rendición de cuentas algorítmica es necesaria para ayudar a los desarrolladores de sistemas algorítmicos y gerentes de empresas a diseñar e implementar sistemas de manera responsable. La rendición de cuentas algorítmica incluye la obligación de informar, explicar o justificar la toma de decisiones algorítmicas, así como la posibilidad de prever y mitigar cualquier impacto social negativo o daño potencial. Los algoritmos automatizados de toma de decisiones ahora se utilizan tanto en el sector privado como en los servicios públicos, lo que sustenta muchos procesos, desde la fijación dinámica de precios hasta las prácticas de empleo y las sentencias penales.

Se parte de la hipótesis que los algoritmos y los datos que utilizan están diseñados o fabricados por personas: siempre hay un ser humano responsable de las decisiones tomadas o informadas por un algoritmo. Por tanto, la frase el algoritmo lo hizo o es el sistema algorítmico, no es una respuesta aceptable en todos los casos. Y suponiendo que el sistema algorítmico haya cometido errores o consecuencias no deseadas, existe en alguna fase la participación de personas (incluso en el *Deep Learning*).

A lo largo de las fases de desarrollo, implementación y puesta en operación de los algoritmos se debe verificar que funcionen con cinco principios básicos. Nicholas Diakopoulos y otros académicos desarrollaron un documento electrónico denominado *Principles for accountable algorithms and a social impact statement for algorithms* en el cual se pueden encontrar estos cinco principios y una guía básica para elaborar una auditoría algorítmica:

“Responsabilidad

Poner a disposición vías de reparación externas visibles para los efectos adversos individuales o sociales de un sistema de decisión algorítmica, y designar un rol interno para la persona responsable de la reparación oportuna de tales problemas.”

“Explicabilidad

Asegúrese de que las decisiones algorítmicas, así como cualquier dato que impulse esas decisiones, se puedan explicar a los usuarios finales y a todo aquel interesado (en términos no técnicos).”

“Exactitud

Identificar, registrar y articular los orígenes del error e incertidumbre en todo el algoritmo y los orígenes de los datos para que se puedan comprender las

implicaciones esperadas y, en el peor de los casos, informar los procedimientos de mitigación que se pueden usar.”

“Auditabilidad

Permitir que terceros interesados analicen, comprendan y revisen el comportamiento del algoritmo mediante la divulgación de información que permita el monitoreo, la verificación o la crítica, incluso mediante el suministro de documentación más detallada.”

“Equidad

Asegúrese de que las decisiones algorítmicas no generan impactos discriminatorios o injustos al comparar diferentes datos (por ejemplo, raza, sexo, etc.).”³¹⁶

Para aquellas auditorías algorítmicas que afectan a la privacidad de las personas o que tiene relación con el impacto de la experimentación científica en seres humanos -que son los dos principios restantes- los autores recomiendan tomar como pautas de revisión las recomendaciones de la OCDE y las reglas de la Oficina de Protección de la Investigación Humana del Departamento de Salud y Servicios Humanos de los Estados Unidos, respectivamente:

- a) Directrices OCDE. - Las directrices de la OCDE sobre protección de la privacidad y flujos transfronterizos de datos personales tienen por objeto ayudar a los responsables de políticas públicas en cada país para la elaboración de marcos de privacidad. Se adoptaron inicialmente en 1980 y se revisaron en 2013. Establecen principios para crear los marcos de protección de la privacidad en los países miembros de la OCDE. La mayor parte de la legislación que existe en materia de privacidad y protección de datos están en consonancia con las Directrices de Privacidad de la OCDE.³¹⁷

³¹⁶ Para conocer más sobre este documento consúltese el Anexo 1 de este ensayo y el sitio: Principles for Accountable Algorithms and a Social Impact Statement for Algorithms: FAT ML. La investigación de Diakopoulos (2020) se centra en el periodismo computacional con proyectos de investigación sobre responsabilidad algorítmica y transparencia, automatización y algoritmos en la producción de noticias; y, redes sociales en contextos de noticias. Es autor del libro *Automating the News: How Algorithms are Rewriting the Media* de Harvard University Press, 2019.

³¹⁷ La protección de la privacidad está regulada por instrumentos internacionales como: 1) Declaración Universal de los Derechos Humanos (1948), el Pacto Internacional de Derechos Civiles y Políticos (1966), el Pacto Internacional de Derechos Económicos, Sociales y Culturales (1966) y la Comisión Americana sobre Derechos Humanos (1969). El principal objetivo es desarrollar e implementar un marco que proteja la privacidad y, a su vez, fomente el uso del entorno digital para la prosperidad económica y social; así como los flujos transfronterizos de datos personales. Para más información consulte las *Directrices de privacidad de la OCDE*, en <http://www.oecd.org/digital/ieconomy/privacy-guidelines.htm>

- b) El Informe Belmont fue escrito por la Comisión Nacional para la Protección de Sujetos Humanos de Investigación Biomédica y Conductual. La Comisión, creada como resultado de la Ley Nacional de Investigación de 1974, se encargó de identificar los principios éticos básicos que deben respetarse para la realización de investigaciones biomédicas y conductuales que involucran a seres humanos y elaborar directrices para garantizar que dicha investigación se lleve a cabo de conformidad con esos principios.³¹⁸

Adicional a las recomendaciones de Diakopoulos y el equipo docente, se debe tomar en cuenta la Recomendación sobre Inteligencia Artificial que emitió la OCDE el 22 de mayo de 2019, la cual incluye dos secciones sustantivas:

- Principios para la administración responsable de una IA confiable: 1. crecimiento inclusivo, desarrollo sostenible y bienestar; 2. valores y equidad centrados en el ser humano; 3. transparencia y explicabilidad; 4. robustez, seguridad y protección; y 5. rendición de cuentas.
- Políticas nacionales y cooperación internacional para una IA confiable: 1. invertir en investigación y desarrollo de IA; 2. fomentar un ecosistema digital para la IA; 3. dar forma a un entorno político propicio para la IA; 4. desarrollar la capacidad humana y prepararse para la transformación del mercado laboral; y 5. la cooperación internacional para la confianza AI.

La Recomendación sobre la Ética de la Inteligencia Artificial, es la primera norma intergubernamental sobre IA adoptada por el Consejo de la OCDE a nivel ministerial a propuesta de su Comité sobre Política de Economía Digital (CDEP). La recomendación tiene como objeto “fomentar la innovación y la confianza en la IA mediante la promoción de la administración responsable de una IA fiable al tiempo que garantiza el respeto de los Derechos Humanos y los valores democráticos. Complementando los estándares existentes de la OCDE en áreas como la privacidad, la gestión de riesgos de seguridad digital y la conducta empresarial

³¹⁸ Para más información consulte el Informe de Belmont así como otras regulaciones, políticas y orientación en temas de manipulación de personas con fines de investigación en: [The Belmont Report | HHS.gov](https://www.hhs.gov/ Belmont-Report/)

responsable, la Recomendación se centra en cuestiones específicas de la IA y establece un estándar que es implementable y suficientemente flexible para resistir la prueba del tiempo en este campo en rápida evolución. En junio de 2019, en la Cumbre de Osaka, los líderes del G20 dieron la bienvenida a los Principios de IA del G20, extraídos de la Recomendación de la OCDE.”³¹⁹

Estos cambios abren el mercado a empresas que están evaluando los riesgos por el uso de algoritmos mediante auditorías especializadas. No hay una fórmula única, cada empresa o persona ofrece o encuentra una estructura de evaluación. Sin embargo, por lo general los problemas que hemos mencionado a lo largo de este apartado son los que más preocupación y más atención demandan. Hay auditorías internas y auditorías externas. A la larga, las legislaciones parecen que tomaran las segundas como un requisito para las empresas que utilizan el ML.

Por ejemplo, Koa Health es un startup que desarrolla terapias digitales y de bienestar para la salud mental y utiliza inteligencia artificial sumada a ciertos principios éticos para preservar la privacidad de los usuarios, evitando posibles sesgos en el diseño de sus aplicaciones. Desde cierto punto de vista, comenta Oliver Smith, director de Estrategia y responsable de ética en Koa, “el objetivo de intentar proteger la privacidad puede colisionar con el de evitar sesgos, ya que para entender que una aplicación está sesgada es necesario saber el género, etnia y algunos otros datos personales de los usuarios. ¿Qué hacer entonces? “No hay una respuesta generalizada, cada organización debe encontrar la suya, pero siempre hay estas fuentes de tensión”, dice Smith. El directivo comentó que es más importante la privacidad y que pueden abordar el sesgo mediante investigaciones indirectas o más específicas. Por ejemplo, en lugar de recopilar datos personales de todos los usuarios, creen que podría ser útil conducir investigaciones con un subgrupo representativo de usuarios del cual sí obtengan datos personales que puedan extrapolar para saber si su *app* está sesgada de alguna forma.”³²⁰

³¹⁹ OCDE, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, p. 3.

³²⁰ Consúltese la entrevista en: <https://www.elespanol.com/invertia/disruptores-innovadores/opinion/20210130/lecciones-practicas-inteligencia-artificial->

Otra empresa involucrada en estos temas, pero desde un enfoque de la Ética en el uso de algoritmos y de fiscalización de su funcionamiento para evitar daños mediante auditorías algorítmicas es *Éticas Research & Consulting*³²¹, quienes evalúan cada caso en particular y lo que conviene realizar, por ejemplo: auditorías de sesgos, auditorías de inspecciones regulatorias, auditorías realizadas en las fases de desarrollo, auditorías realizadas ya en ambiente de producción.

Las empresas están incorporando algoritmos de IA en sus procesos - comerciales, financieros, de recursos humanos, antilavado de dinero, etcétera- sin embargo, a medida que van incorporando estos algoritmos también aumentan los riesgos a los que se expone la empresa, sobre todo, cuando el usuario o cliente se ve afectado en su persona, patrimonio, imagen, planes de vida, al no funcionar correctamente en las predicciones o cálculos que realizan. Cuando estos errores son a gran escala, es decir, se vuelven una constante que se replica en varios casos similares y durante un período muy corto es cuando la empresa busca corregir el algoritmo y para ello se recomienda a experto en auditorías algorítmicas.

Una auditoría algorítmica está pensada para entender cómo están procesándose los datos por el algoritmo y si hay el elemento de explicabilidad. Al operar la IA basada en Machine Learning – Deep Learning con algoritmos de caja negra, los resultados no son explicables, pero, los expertos han buscado conocer el proceso de toma de decisiones del algoritmo y qué tan preciso o qué tan eficiente es para los casos que se le confían. Un algoritmo debe ser confiable y explicable. Al interior de la empresa cada área ya tiene una carga de responsabilidad que muchas veces es suficiente para mantener al equipo ocupado y saturado. Por tanto, se recomienda contratar auditores externos que colaboren con una visión imparcial y objetiva sobre el uso de los algoritmos y los riesgos que se pueden generar con el fin de mitigarlos; de crear un plan de riesgos en las áreas de la empresa que estén utilizando algoritmos. El mayor riesgo de una empresa es el riesgo reputacional.

[etica/554574546_13.html?utm_term=Autofeed&utm_campaign=Echobox&utm_medium=Social&utm_source=LinkedIn](https://www.eticasconsulting.com/etica/554574546_13.html?utm_term=Autofeed&utm_campaign=Echobox&utm_medium=Social&utm_source=LinkedIn)

³²¹ Para conocer más de *Éticas Research & Consulting* puede consultarse su página en: <https://www.eticasconsulting.com/>

Una auditoría algorítmica da a la empresa y al gobierno la posibilidad de la mejora continua en sus procesos y, sobre todo, la posibilidad de detectar y corregir permanentemente cualquier violación a Derechos Humanos de las personas que utilicen sus algoritmos. Por ejemplo, cuando tienen una idea para echar andar un servicio o vender un producto basado en inteligencia artificial, las empresas hacen pruebas previas con un grupo reducido de usuarios o clientes. Pero cuando se abre al mercado global, ese grupo reducido no va a representar la misma clase de usuarios o clientes y aquí pudieran aparecer fallas o sesgos en el uso del algoritmo. En otras ocasiones, puede estar funcionando bien a nivel macro, pero debido a las circunstancias del mercado o a externalidades como el COVID-19 pudiera ser necesario una revisión al algoritmo para validar que no esté generando una afectación a la empresa, usuario, cliente o a un grupo social (sexo, raza, edad, preferencia sexual).

El tema de la regulación algorítmica ya sea a través del Derecho o de la Ética requiere de principios, reglas, estándares de medición y buenas prácticas sobre transparencia, privacidad, explicabilidad, responsabilidad, equidad, sesgos, etcétera. Empresas como LI²FE están creando informes de impacto ético, auditorías éticas, diseño e implementación de estrategias éticas, códigos de buenas prácticas para el uso de IA, gestión de datos y privacidad; Ética Digital y planes de Gobernanza Ética.³²²

El campo de la auditoría ética algorítmica es un terreno que va a cobrar gran auge en los próximos años y se está convirtiendo en un requisito de operación o utilización de cualquier algoritmo de IA tanto en la Unión Europea como en los Estados Unidos de América. Estas auditorías son necesarias para detectar los posibles riesgos que genera la IA y hacerla interpretable, explicable, transparente, predecible y controlable por el ser humano. La auditoría también permite detectar el grado o tramo de responsabilidad algorítmica.

El éxito de un modelo algorítmico está, en parte, en la interacción entre el humano y la máquina y en el uso que las personas hacen de este modelo. Al

³²² Para conocer más de esta empresa puede consultarse su página en: [LI²FE \(site123.me\)](https://site123.me)

momento de hacer una auditoría algorítmica se debe proporcionar información clara sobre qué mide el algoritmo, para qué puede contribuir su uso y cómo complementa sus actividades o funciones. La persona que utilice el algoritmo debe saber que siempre, al final, hay responsabilidades y sesgos, cada etapa de un proyecto en el que se involucra ML puede tener sesgos, cada decisión puede tener un sesgo. Las auditorías algorítmicas para detectar sesgos deben tomar en cuenta lo anterior. Las auditorías algorítmicas serán un ingrediente necesario e indispensable para la empresa no solo para evitar responsabilidades, sino también para mejorar y corregir, aprender y construir modelos que de otra manera sería imposible de hacer. El enfoque basado en riesgos contendrá todos los ingredientes necesarios para mitigar los riesgos para la empresa (internos y externos). La siguiente tabla muestra algunas herramientas para la realización de auditorías con un enfoque basado en riesgos:

Herramientas para la realización de una auditoría algorítmica con EBR	
Herramienta	Sitio web
Google	Tools – Google AI https://www.tensorflow.org/hub
H2O	https://www.h2o.ai/
IBM.	http://aif360.mybluemix.net/ https://developer.ibm.com/es/
AWS	https://aws.amazon.com/es/sagemaker/clarify/ https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-processing-job-run.html
Aequitas	https://aif360.mybluemix.net/

Fuente: Elaboración propia con base en búsquedas en Google.

Una interesante propuesta para elaborar una auditoría algorítmica basada en riesgos es la Guía de Supervisión Modelo sobre gestión de riesgos, publicada por el *Board of Governors of the Federal Reserve System*. El término modelo -dice la Guía- se refiere a un método cuantitativo, sistema o enfoque que aplica teorías, técnicas y supuestos estadísticos, económicos, financieros o matemáticos para procesar datos de entrada en estimaciones cuantitativas. Los modelos que cumplan con esta definición podrían utilizarse para analizar estrategias de negocio, informar las decisiones empresariales, identificar y medir riesgos, valorar exposiciones,

instrumentos o posiciones, realizar pruebas de estrés, evaluar la adecuación del capital, gestionar los activos de los clientes, medir el cumplimiento de los límites internos, mantener el aparato de control formal del banco o cumplir con los requisitos de informes financieros o reglamentarios y emitir divulgaciones públicas. Esto se puede llevar al campo de las auditorías algorítmicas adaptando lo correspondiente a cada sector económico o productivo.³²³

Para finalizar, esta sección es de gran relevancia conocer el trabajo de Alessandro Mantelero y M. S. Esposito.³²⁴ Estos autores proponen una metodología para la Evaluación del Impacto en Derechos Humanos (EIDH), que, como dijimos anteriormente, sale del enfoque centrado en los datos o del enfoque sentado en la Ética y lo trasladan hacia el ámbito de protección de todos los Derechos Humanos, lo cual representa una protección que va del individuo hacia determinado grupo o grupos de personas, poblaciones y la humanidad en su conjunto. Reconocer que los Derechos Humanos están presentes en todo y protegen a todos juega un papel clave para la regulación de la IA basada en Machine Learning – Deep Learning. La metodología propuesta fue utilizada en estudios de casos concretos para demostrar su viabilidad y eficacia, dejando el mero debate teórico para pasar a una implementación real y específica en el campo de las aplicaciones de IA.

El debate -dicen los autores- se ha caracterizado por vincular la problemática de la IA 1) hacia el uso de los datos, haciendo énfasis en la dimensión Ética del uso de algoritmos y, en el ámbito legal, se ha hablado mucho de la 2) discriminación por sesgos y su protección. Sin embargo, la sociedad impulsada por la IA muestra una variedad de impactos potenciales sobre los derechos y libertades individuales y colectivas, lo que sugiere, ampliar el campo de protección hacia instrumentos legales (no reemplazables por pautas éticas), y, además, la necesidad de un análisis del impacto de la IA en los Derechos Humanos; y, la creación y aplicación de un

³²³ Cfr. Board of Governors of the Federal Reserve System (2011). “The Fed -Supervisory Letter SR 11-7 on Guidance on Model Risk Management”. <https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>

³²⁴ Mantelero, Alessandro y Esposito, M.S. *An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems*, Computer Law & Security Review, 2021

modelo de evaluación de impacto específico de la IA. Al crear un enfoque metodológico para la evaluación de impacto basado en tratados internacionales, se produce un modelo que puede usarse más fácilmente en diferentes partes del mundo (basado en la naturaleza universal de los Derechos Humanos).

Los autores están convencidos de que el enfoque ético ha sido superado y que se requiere retomar el camino legal, sugiriendo un papel complementario entre la Ética y la ley y enfatizando que solo los Derechos Humanos pueden proporcionar una referencia uniforme para regular la IA en diferentes contextos culturales, éticos y legales. Para comprobar su propuesta eligieron como caso de estudio a la Unión Europea y al Reglamento General sobre Protección de Datos. El objeto es encontrar y extraer la lista de derechos y libertades potencialmente impactados basada en evidencia concreta en contraposición a casos e hipótesis teóricos. También se propone encontrar las limitaciones en las regulaciones de protección de datos para abordar los problemas de IA y la necesidad de un enfoque más personalizado.

Una vez acotado el marco teórico al reglamento general sobre protección de datos y reunida la evidencia empírica, se presenta el producto de la investigación, el desarrollo y la prueba de un modelo de evaluación de impacto en los Derechos Humanos centrado en la inteligencia artificial. El análisis contrafactual de la documentación disponible permitió probar el modelo y mostrar sus resultados o efectos en esos contextos.

El análisis hecho por los autores demuestra que las afectaciones y los problemas relacionados con el uso de aplicaciones de IA no se circunscriben únicamente al tema de los sesgos y la discriminación, sino que tienen un impacto más amplio en varios Derechos Humanos. Esta hipótesis es compartida por nosotros, y creemos que la metodología para la evaluación integral de Derechos Humanos, necesita incorporarse como un ingrediente del ecosistema de protección social para el uso de IA, que deben utilizarse adicional a la generación de políticas públicas e instituciones, presupuestos y demás insumos que se requiere para la protección de la humanidad.

Pensemos tan solo que en unos cuantos años, muchas ciudades serán ciudades inteligentes, tendrán planes regionales de movilidad inteligente, programas de predicción de delitos inteligentes, ciberjusticia ampliada a muchas materias, entre otros temas. Adicionalmente, encontramos a las empresas y usuarios de estas tecnologías que están utilizando productos y servicios distribuidos globalmente (por ejemplo, asistentes virtuales de IA, vehículos autónomos, software de contratación basado en IA, entre otros) que no se centran en una comunidad territorial determinada.

La herramienta también fue pensada para acompañar a la IA desde su creación u origen, su entrenamiento, etapa de producción hasta que concluya el ciclo de vida del sistema inteligente, proporcionando evidencia específica, medible y comparable sobre los impactos potenciales, su probabilidad, extensión, y facilitando la comparación entre opciones de diseño alternativas para el diseño de IA, basado en la evaluación y mitigación de riesgos. En este sentido, el modelo se convierte en una herramienta de gestión de Derechos Humanos, que proporciona evidencia clara para un desarrollo orientado a los Derechos Humanos de los productos y servicios de IA. El modelo fue pensado para su uso en la empresa y por las autoridades supervisoras u órganos de auditoría para el monitoreo y la gestión de riesgos en relación con el impacto del uso de datos en los que se vean involucrados los Derechos Humanos.

Una vez que la normatividad de la Unión Europea se actualice y se vuelva obligatoria, la MEIDH se convierta en una herramienta legal, necesaria o, mejor dicho, indispensable, tal como es hoy el enfoque basado en riesgos de las entidades financieras. La futura regulación de la IA traerá esa y otras obligaciones legales específicas y, por obvias razones, sanciones por incumplimiento en relación con la evaluación y gestión de riesgos producidos por sistemas de IA que afecten Derechos Humanos y permita ubicar la responsabilidad de las personas (en caso de existir y poder probarla).

Así como surgió el Grupo de Acción Financiera Internacional (GAFI), tendrá que crearse algo similar para la regulación de la IA en el mundo y se volverá

obligatorio para las empresas contar con un oficial de cumplimiento en materia de Derechos Humanos que tenga que aplicar y supervisar el cumplimiento de su manual de políticas y procedimientos, el código de Ética y el código de conducta, la utilización de un sistema automatizado de monitoreo, capacitación permanente y obligatoria para que toda la empresa entienda el tema y su problemática; la creación de nuevas estructuras, responsabilidades e indicadores, entre otros. Todo para agilizar o facilitar un enfoque de la IA respetuoso de Derechos Humanos.

Capítulo III. IBM Watson y el caso *Watson Health*

Capítulo III. IBM Watson y el caso *Watson Health*

Esta sección tiene como *objetivo particular* examinar, por un lado, a *IBM Watson*, sus *características y funciones*, así como algunos casos ejemplificativos de uso y el parámetro de evaluación. Y, por el otro, entender El caso *Watson Health* en el contexto de la *Healthcare technology*, sus características y la problemática de su uso.

Actualmente, la IA basada en Machine Learning – Deep Learning y las nuevas tecnologías que la utilizan, están cambiando de manera sorprendente la vida de todos los seres humanos. Watson de IBM es un ejemplo de este tipo de tecnologías que contribuye a la digitalización de la vida de las personas y el inicio de la Cuarta revolución industrial. Para efectos de esta investigación tomamos como base la definición de Klaus Schwab, fundador y director general del Foro Económico Mundial:

“La primera revolución industrial abarcó desde aproximadamente 1760 hasta más o menos 1840. Desencadenada por la construcción del ferrocarril y la invención del motor de vapor, marcó el comienzo de la producción mecánica. La segunda revolución industrial, entre finales del siglo XIX y principios del XX, hizo posible la producción en masa, fomentada por el advenimiento de la electricidad y la cadena de montaje. La tercera revolución industrial se inició en la década de 1960. Generalmente se le conoce como la revolución digital o del ordenador, porque fue catalizada por el desarrollo de los semiconductores, la computación mediante servidores tipo “mainframe” (en los años setenta), la informática personal (décadas de 1970 y 1980) e internet (década de 1990) ... creo que ahora estamos en una cuarta revolución industrial. Esta comenzó a partir de este siglo y se basa en la revolución digital. Se caracteriza por un internet más ubicuo y móvil, por sensores más pequeños y potentes que son cada vez más baratos, y por la inteligencia artificial y el aprendizaje de la máquina... En Alemania se debate sobre la “industria 4.0” un término acuñado en la Feria de Hannover de 2011 para describir cómo esta revolucionará la organización de las cadenas de valor globales. Mediante la creación de “fábricas inteligentes”, la cuarta revolución industrial genera un mundo en el que sistemas de fabricación virtuales y físicos cooperarán entre sí de una manera flexible en todo el planeta. Esto permite la absoluta personalización de los productos y la creación de nuevos modelos de operación... Al mismo tiempo, se producen oleadas de más avances... que van desde la secuenciación genética

hasta la nanotecnología, y de las energías renovables a la computación cuántica...”³²⁵

El sistema Watson, como otros de este tipo³²⁶, representan una de las herramientas que más valor aportarán a la Cuarta revolución industrial y consideramos necesario ubicar determinados momentos y conceptos que son el origen de todo lo que se está desarrollando en el siglo XXI.

Los ingredientes de la computación cognitiva que utiliza Watson para su funcionamiento contribuyeron a los nuevos avances tecnológicos de la cuarta revolución industrial.³²⁷ El término computación cognitiva se usa generalmente para describir sistemas de inteligencia artificial que tienen como objeto simular las capacidades de la mente humana. La cognición humana implica el análisis de su entorno, el contexto y la intención, entre muchas otras variables, todo esto en tiempo real, y permite a las personas resolver problemas. A decir de los investigadores de estos temas, se requieren varias tecnologías de inteligencia artificial para que un sistema informático construya modelos cognitivos que imiten los procesos de la mente humana, incluido el aprendizaje automático, el aprendizaje profundo, las redes neuronales, la PNL y el análisis de sentimientos.³²⁸

³²⁵ Schwab. Klaus, *La cuarta revolución industrial*, Penguin Random House, México, 2017, pp. 20 y 21.

³²⁶ Para conocer, de manera general, algunos aspectos de los 4 principales proveedores de esta tecnología, consúltese: IBM (Watson – The debater) *The Debater* https://www.youtube.com/watch?v=koec1nnePAM&feature=emb_logo; Google (DeepMind –ALPHA GO ZERO) *AlphaGo Zero: Discovering new knowledge* <https://www.youtube.com/watch?v=WXHFqTvfFSw&t=1s>; Amazon (AWS) *What is AWS?* https://www.youtube.com/watch?v=a9_D53WsUs Microsoft AZURE *How does Microsoft Azure work?* <https://www.youtube.com/watch?v=KXkBZCe699A> Whiting, Kate, “Estas son las 10 principales habilidades laborales del futuro - y el tiempo que lleva aprenderlas”, consúltese en: <https://es.weforum.org/agenda/2020/10/estas-son-las-10-principales-habilidades-laborales-del-futuro-y-el-tiempo-que-lleva-aprenderlas/?fbclid=IwAR1vCiQhob-R0ZikoCkHlbuctuiNQULxoHeBNsXNCYZFJ3-gjPNbEKktXWw> [15 de julio de 2022]

³²⁷ Thomas M. Siebel, en su libro *Digital Transformation. Survive and Thrive in an Era of Mass Extinction* (2019, RossettaBooks, New York) ubica 4 tecnologías que están cambiando el mundo del ser humano: 1) Computación en la nube; 2) Big Data; 3) El renacimiento de la Inteligencia Artificial (a través del *Machine Learning* y el *Deep Learning*); y, 4) Internet de las Cosas.

³²⁸ Cfr. Dharmendra S. Modha, Rajagopal Ananthanarayanan, Steven K. Esser, Anthony Ndirango, Anthony J. Sherbondy, and Raghavendra Singh. 2011. “Cognitive computing”. *Commun. ACM* 54, 8, agosto de 2011, pp. 62 a 71. DOI: <https://doi.org/10.1145/1978542.1978559> [15 de julio de 2022]

Antes de entender las características y funciones de Watson así como los casos y su evaluación, debemos dimensionar el papel que ha tenido IBM como empresa de tecnología. IBM tiene sus orígenes el 16 de junio de 1911 cuando Charles R. Flint (1850-1934), organiza la fusión de International Time Recording Company, Computing Scale Company y Tabulating Machine Company para formar la *Computing-Tabulating-Recording Company* (CTR). Para 1924, la Computing-Tabulating-Recording Co. se transforma en *International Business Machines Corporation* (IBM). Años antes de su transformación, Thomas J. Watson, Sr., se une a Computing Tabulating Recording Co. (CTR) como gerente general (1914) y, al año siguiente, Watson es elegido presidente de la empresa. No se puede entender la mística de la compañía sin conocer el liderazgo comercial que le dio Thomas J. Watson, Sr., y su fórmula de los cinco pasos hacia el conocimiento: “Leer, escuchar, discutir, observar, pensar” (think).³²⁹ El sistema Watson que se analiza en este texto, lleva el apellido de este directivo, quien supervisó el crecimiento de la empresa hasta convertirla en una multinacional.

En 1944, IBM presenta su primera computadora a gran escala, la *Automatic Sequence Controlled Calculator* (ASCC o Calculadora Automática de Secuencia Controlada), conocida como Mark I. La ASCC es la primera máquina que ejecuta cálculos largos automáticamente, utilizaba relés electromecánicos para realizar sumas, multiplicaciones y divisiones.

En 1952, la empresa presentó el IBM 701, una computadora a base de tubos de vacío. Los tubos eran más rápidos, pequeños y fáciles de reemplazar, a diferencia de los interruptores electromecánicos del equipo Mark I (1944). El 701 podía ejecutar 17,000 instrucciones por segundo, los tubos de vacío movieron rápidamente las computadoras a aplicaciones comerciales como facturación, nómina y control de inventario iniciando un proceso de equipamiento tecnológico en las empresas. Para 1959, los transistores reemplazarían a los tubos de vacío.³³⁰

³²⁹ Consúltese: *IBM Archives Valuable resources on IBM's history*, en el sitio: <https://www.ibm.com/ibm/history/index.html>

³³⁰ *Idem*.

En 1957, IBM crea el lenguaje de programación FORTRAN (*FORmula TRANslation*). FORTRAN se convierte en el lenguaje informático más utilizado para trabajos técnicos en todo el mundo. Servirá para programar los sistemas expertos.³³¹

En 1967, Frank Rosenblatt construye el *Mark 1 Perceptron*, la primera computadora basada en una red neuronal que aprendió mediante el método de aproximaciones sucesivas (ensayo-error); solo un año después, Marvin Minsky y Seymour Papert publican un libro titulado *Perceptrons*, que se convierte tanto en el trabajo histórico sobre redes neuronales como, al menos por un tiempo, en un argumento en contra de futuros proyectos de investigación de redes neuronales.³³²

En 1971, se crea la primera aplicación operativa de reconocimiento de voz que permite a los ingenieros hablar y recibir respuestas habladas desde una computadora que puede reconocer alrededor de 5,000 palabras. IBM también desarrolla una terminal experimental que imprime respuestas de computadora en Braille para ciegos.³³³

En 1980, presenta las redes neuronales con retropropagación (algoritmos para entrenar la red) que hoy en día se utilizan ampliamente en aplicaciones de inteligencia artificial. En 1997, Deep Blue de IBM vence al entonces campeón mundial de ajedrez Garry Kasparov, en una partida de ajedrez (y en la revancha).³³⁴

En 2007, se crea IBM Watson, un sistema diseñado para hacer preguntas y generar respuestas profundas, análisis de información y la comprensión de la computadora del lenguaje natural. Watson se convirtió en la primera computadora cognitiva, combinando el aprendizaje automático, las redes neuronales profundas e inteligencia artificial tradicional basada en sistemas expertos decisionales. En 2011, IBM Watson vence a los campeones Ken Jennings y Brad Rutter en Jeopardy.

³³¹ *Idem.*

³³² *Idem.*

³³³ *Idem.*

³³⁴ *Idem.*

En 2019, presentó el IBM Q System One, el primer ordenador cuántico para uso comercial. En el mismo se combina tanto la computación cuántica como tradicional para ofrecer un sistema de 20 qubits para su utilización en investigaciones y grandes cálculos.³³⁵ En 2020, IBM presenta Project Debater³³⁶. Watson Rob Thomas, Gerente General de IBM Data and AI, anunció que los productos de inteligencia artificial centrales de Watson incluirán capacidades de procesamiento de lenguaje natural (PNL). Las nuevas tecnologías representan la primera comercialización de capacidades clave de PNL que provienen de Project Debater de IBM Research. Sus características son:

- Análisis de sentimientos avanzado: Watson puede identificar y comprender mejor los esquemas de palabras complicados, a veces coloquiales, como modismos y cambios de sentimiento, como apenas útil, y sus significados. Esta característica también incluye una nueva tecnología de clasificación que permitirá a los clientes crear modelos de PNL que clasificarán más fácilmente el lenguaje comercial específico de la industria.
- Recapitulación: Watson puede analizar una variedad de fuentes y proporcionar un resumen o breve de las ideas y la información que contiene.
- Agrupación: las nuevas técnicas de agrupación de temas brindan a los expertos en la materia una herramienta para agrupar los datos entrantes en temas para que puedan examinar los temas y ver grupos de información relacionada. Se pueden ajustar a los temas para reflejar el lenguaje de empresas o industrias específicas, como seguros, atención médica y manufactura.

El objetivo de Project Debater, narra Dakshi Agrawal, IBM Fellow and Chief Architect for AI, es ayudar a las personas a construir argumentos persuasivos y tomar decisiones bien informadas, independientemente de sus prejuicios

³³⁵ *Idem.*

³³⁶ Puede consultarse: IBM, What happens when AI stops playing games? <https://www.youtube.com/watch?v=7pHaNMdWGsk>

personales.³³⁷ Entremos ahora a la descripción de 1) las características y funciones de IBM Watson y a los 2) casos prácticos de su uso, así como su evaluación dentro del cuadrante de sistemas inteligentes de Gartner.³³⁸

1) IBM Watson

Watson utiliza los algoritmos de *Deep Learning* para crear una red neuronal artificial que le permite aprender continuamente, determinar si las decisiones fueron correctas, mejorar constantemente la calidad y la precisión de los resultados. Esto, a su vez, le permite aprender de datos no estructurados como fotos, videos y archivos de audio. El aprendizaje profundo también da las capacidades de comprensión del lenguaje natural, permitiéndole aprender deconstruyendo oraciones, luego analizarlas e identificar conceptos y sus relaciones con esas oraciones. Una vez que comprende cómo encaja todo, puede resolver el contexto y la intención de lo que se transmite.

Los modelos tradicionales de IA requerían de muchos datos y gran potencia de los equipos de cómputo, tomando semanas o meses para su consolidación. Watson, lo hace de forma rápida y sencilla, gracias a una técnica llamada *Transfer Learning* (Transferir el aprendizaje es aprender a aprender). Es lo que permite a Watson aprender más rápido con menos tiempo, por lo que no necesita ser entrenado desde cero. También se puede alimentar con conocimiento previo para acelerar aún más los resultados. Las siguientes imágenes nos dan detalles de cómo funciona Watson.

³³⁷Agrawal, Dakshi. *La IA aprende el lenguaje de los negocios: entre bastidores en el evento de innovación de PNL de IBM*, IBM, publicado el 17 de marzo de 2020 en el sitio: <https://www.ibm.com/blogs/watson/2020/03/ai-learns-the-language-of-business-behind-the-scenes-at-ibms-nlp-innovation-event/> [15 de julio de 2022]

³³⁸ Para otros datos relacionados con la historia de IBM consúltase el Anexo 4 “Historia de IBM” de esta sección.

¿Qué es Watson? ¿Cómo funciona?

¿Qué es Watson? ¿Cómo funciona?

Watson es una plataforma de inteligencia artificial empresarial.

Watson aprende e interactúa como lo hacemos los humanos.

¿Cómo funciona Watson?

- Entiende**
Los sistemas cognitivos entienden imágenes, lenguaje y otros datos no estructurados, igual que los humanos.
- Razona**
Pueden razonar, entender conceptos que hay detrás, formular hipótesis, inferir y extraer ideas.
- Aprende**
Con cada dato o interacción afinan su experiencia y nunca dejan de aprender.
- Interactúa**
Con capacidades para ver, hablar y escuchar, los sistemas cognitivos interactúan con los humanos de manera natural.

6

Fuente: Jorge Alberto Paz, exposición en el IBM Educathon, 2020.

Portal Watson, IBM (julio, 2022)

My projects / Jose Luis Hernandez S

Overview Assets

Jose Luis Hernandez S
Last Updated: Jun 10, 2020
[Readme](#)

Overview
Date created
Jun 10, 2020
Description
Watson VR service
Storage
0 Byte used
Cloud Object Storage

Choose asset type

Available asset types

- Data
- Connection
- Connected data
- AutoAI experiment
- Notebook
- Dashboard
- Visual Recognition ...
- Natural Language CL...
- Watson Machine Lea...
- Deep learning experi...
- Modeler flow
- Data Refinery flow
- Streams flow
- Decision Optimizatio... **NEW**

Fuente: <https://cloud.ibm.com/> [15 de julio de 2022]

En cualquier industria o servicio, Watson puede eficientar las actividades de la organización. Un ejemplo que nos dan los expertos de IBM es pensar en eficientar las actividades de una compañía de seguros de automóviles. El aprendizaje de transferencia de Watson funciona con tres capas: La capa inferior está compuesta

de conocimiento general fuera de la caja (Wikipedia o sitios de ese tipo en internet). Esta capa le dice a Watson qué es un automóvil o un semáforo. La capa intermedia está pre-empaquetada con conocimiento adaptado a industria específica, es decir, el sector de los seguros de automóviles, coberturas, beneficiarios. La capa superior es donde el aprendizaje específico del cliente tiene lugar, personalizando a Watson para las necesidades comerciales únicas de la empresa y vinculándolas con las reglas del sector y el conocimiento general.

A decir de los expertos de IBM, *Transfer Learning* hace que Watson sea considerado como la mejor opción de IA para la empresa ya que le permite aprender más rápido de conjuntos de datos más pequeños, al tiempo que protege sus datos personales e ideas. Watson está diseñado para garantizar que los datos solo fluyan de abajo hacia arriba, para que pueda aprovechar al máximo el conocimiento compartido en las dos capas inferiores, mientras conserva la propiedad de su capa superior. En síntesis, podemos decir que con Watson se puede aprender de datos no estructurados y mejorar constantemente la calidad y la precisión de los datos estructurados.

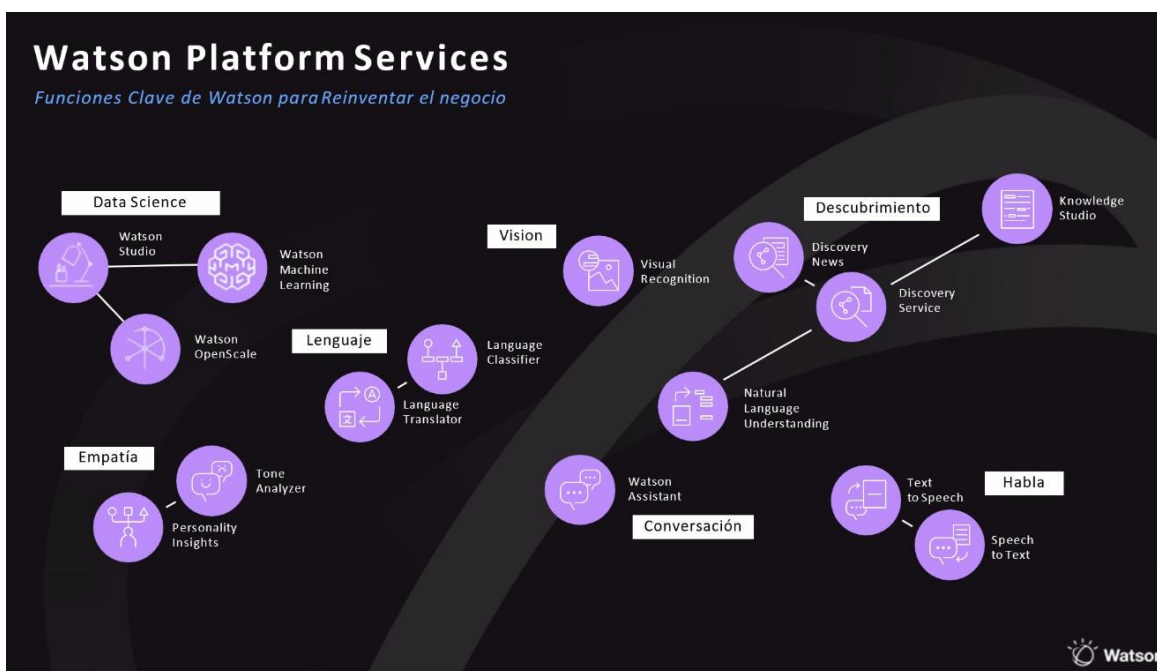
A) Características y funciones

Watson se ha convertido en una gama de APIs avanzadas³³⁹, herramientas especializadas y aplicaciones de software de servicio creadas para las empresas al albergar en la nube híbrida de IBM. Esto significa que Watson está diseñado para atender cualquier tipo de problema o caso -complejo o sencillo- en todas partes del

³³⁹ Una API es “significa interfaz de programación de aplicaciones (Application Program Interface). Las API permiten que sus productos y servicios se comuniquen con otros, sin necesidad de saber cómo están implementados. Esto simplifica el desarrollo de las aplicaciones y permite ahorrar tiempo y dinero. Las API le otorgan flexibilidad; simplifican el diseño, la administración y el uso de las aplicaciones, y proporcionan oportunidades de innovación, lo cual es ideal al momento de diseñar herramientas y productos nuevos (o de gestionar los actuales). Las API son un medio simplificado para conectar su propia infraestructura a través del desarrollo de aplicaciones nativas de la nube híbrida, pero también le permiten compartir sus datos con clientes y otros usuarios externos. Las API públicas representan un valor comercial único porque simplifican y amplían la forma en que se conecta con sus partners y, además, pueden rentabilizar sus datos (un ejemplo conocido es la API de Google Maps).” Definición tomada de: RED HAT, ¿QUÉ ES UNA API? Qué son las API y para qué sirven, en <https://www.redhat.com/es/topics/api/what-are-application-programming-interfaces>

mundo que se tenga acceso a la nube de IBM. IBM garantiza un acceso sin problemas a la información que se necesita para tomar las decisiones correctas. La inteligencia artificial puede ser vista como una escalera o serie de pasos que se deben seguir para el procesamiento de la información y que IBM desarrolla en Watson para que el usuario pueda recolectar, organizar, analizar, integrar e implementar los datos que ingresa.³⁴⁰ La siguiente lámina nos muestra las funciones clave de Watson.

Funciones clave de Watson

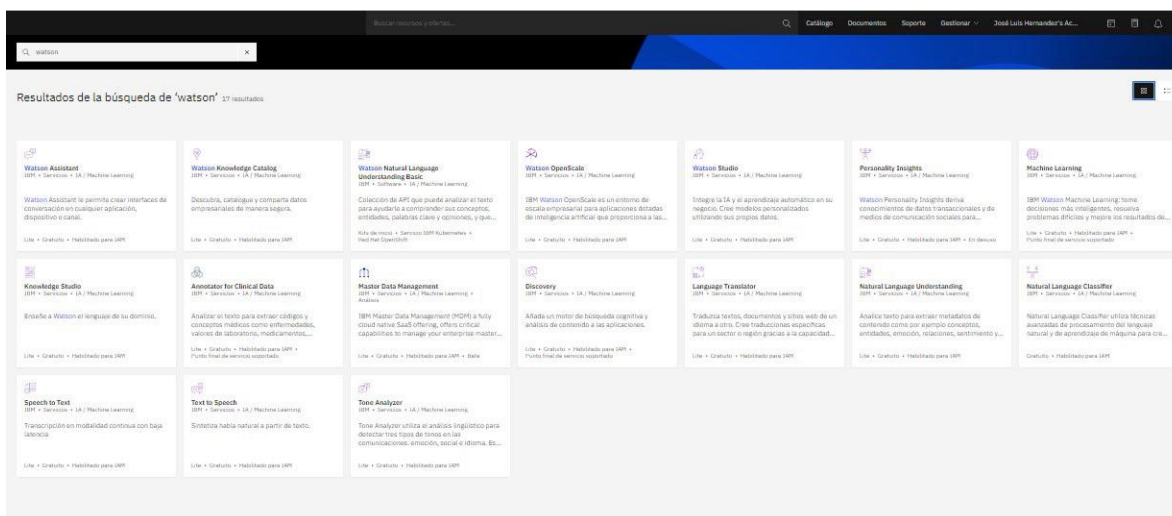


Fuente: Jorge Alberto Paz, exposición en el IBM Educathon 2020.

³⁴⁰ Los expertos de IBM lo definen de la siguiente manera: “Recolectar: simplificar los datos y hacerlos accesibles. Recolectar datos —de todo tipo, sin importar donde residan— y aportar flexibilidad frente a los siempre cambiantes orígenes de los datos. Organizar: crear una base de estudios analíticos listos para la empresa. Organizar todos los datos en una base confiable lista para la empresa con integración de gobernanza, protección y cumplimiento. Analizar: crear y escalar la IA con confianza y transparencia. Analizar los datos de forma más inteligente y beneficiarse con los modelos de IA que empoderan a las organizaciones para ganar nuevos conocimientos y tomar decisiones mejores y más inteligentes. Integrar: operativizar la IA en toda la empresa. Implementar la IA en la empresa en múltiples departamentos y dentro de varios procesos, valiéndose de predicciones, automatización y optimización.” Cfr. Odell, Steve y Narvil Jasper. Estudio “Nueva tecnología: proyección del impacto económico total (*Total Economic Impact™*)”, Forrester-IBM, febrero de 2020.

Watson, como dijimos previamente, se convirtió en la primera computadora cognitiva, combinando el aprendizaje automático, a redes neuronales convolucionales profundas y las metodologías tradicionales de los sistemas expertos. Las siguientes imágenes fueron tomadas de la nube de IBM en abril de 2021:

Funciones de Watson en la nube de IBM -pantalla principal- (julio, 2022) -parte 1-



Fuente: <https://cloud.ibm.com/>

Funciones de Watson en la nube de IBM -pantalla principal- (julio, 2022) -parte 2-

 Watson Assistant IBM • Servicios • IA / Machine Learning Watson Assistant le permite crear interfaces de conversación en cualquier aplicación, dispositivo o canal. Lite • Gratuito • Habilitado para IAM	 Watson Knowledge Catalog IBM • Servicios • IA / Machine Learning Descubra, catalogue y comparta datos empresariales de manera segura. Lite • Gratuito • Habilitado para IAM	 Watson Natural Language Understanding Basic IBM • Software • IA / Machine Learning Colección de API que puede analizar el texto para ayudarlo a comprender sus conceptos, entidades, palabras clave y opiniones, y que... Kits de inicio • Servicio IBM Kubernetes • Red Hat OpenShift	 Watson OpenScale IBM • Servicios • IA / Machine Learning IBM Watson OpenScale es un entorno de escala empresarial para aplicaciones dotadas de inteligencia artificial que proporciona a las... Lite • Gratuito • Habilitado para IAM
 Watson Studio IBM • Servicios • IA / Machine Learning Integre la IA y el aprendizaje automático en su negocio. Cree modelos personalizados utilizando sus propios datos. Lite • Gratuito • Habilitado para IAM	 Personality Insights IBM • Servicios • IA / Machine Learning Watson Personality Insights deriva conocimientos de datos transaccionales y de medios de comunicación sociales para... Lite • Gratuito • Habilitado para IAM • En desuso	 Machine Learning IBM • Servicios • IA / Machine Learning IBM Watson Machine Learning: tome decisiones más inteligentes, resuelva problemas difíciles y mejore los resultados de... Lite • Gratuito • Habilitado para IAM • Punto final de servicio soportado	 Knowledge Studio IBM • Servicios • IA / Machine Learning Enseñe a Watson el lenguaje de su dominio. Lite • Gratuito • Habilitado para IAM
 Annotator for Clinical Data IBM • Servicios • IA / Machine Learning Analizar el texto para extraer códigos y conceptos médicos como enfermedades, valores de laboratorio, medicamentos,... Lite • Gratuito • Habilitado para IAM • Punto final de servicio soportado	 Master Data Management IBM • Servicios • IA / Machine Learning • Análisis IBM Master Data Management (MDM) a fully cloud native SaaS offering, offers critical capabilities to manage your enterprise master... Lite • Gratuito • Habilitado para IAM • Beta	 Discovery IBM • Servicios • IA / Machine Learning Añada un motor de búsqueda cognitiva y análisis de contenido a las aplicaciones. Lite • Gratuito • Habilitado para IAM • Punto final de servicio soportado	 Language Translator IBM • Servicios • IA / Machine Learning Traduzca textos, documentos y sitios web de un idioma a otro. Cree traducciones específicas para un sector o región gracias a la capacidad... Lite • Gratuito • Habilitado para IAM
 Natural Language Understanding IBM • Servicios • IA / Machine Learning Analice texto para extraer metadatos de contenido como por ejemplo conceptos, entidades, emoción, relaciones, sentimiento y... Lite • Gratuito • Habilitado para IAM	 Natural Language Classifier IBM • Servicios • IA / Machine Learning Natural Language Classifier utiliza técnicas avanzadas de procesamiento del lenguaje natural y de aprendizaje de máquina para cre... Gratuito • Habilitado para IAM	 Speech to Text IBM • Servicios • IA / Machine Learning Transcripción en modalidad continua con baja latencia Lite • Gratuito • Habilitado para IAM	 Text to Speech IBM • Servicios • IA / Machine Learning Sintetiza habla natural a partir de texto. Lite • Gratuito • Habilitado para IAM
 Tone Analyzer IBM • Servicios • IA / Machine Learning Tone Analyzer utiliza el análisis lingüístico para detectar tres tipos de tonos en las comunicaciones: emoción, social e idioma. Es... Lite • Gratuito • Habilitado para IAM			

Fuente: <https://cloud.ibm.com/>

A continuación, se hará una explicación detallada de cada una de las funciones de Watson:

- a. *Watson Assistant* (Conversacional): Sirve para crear *chatbots* o asistentes virtuales en dispositivos móviles, plataformas de mensajería e incluso robots. IBM Watson proporciona diferentes APIs para crear asistentes virtuales inteligentes. Recordemos que en sus inicios Watson fue pensado como una máquina para competir y ganar juegos (Jeopardy). El asistente de Watson y los servicios de agente de voz no solo sirven para crear un *chatbot* que pueda comprender las intenciones de los usuarios. La mayoría de los *chatbots* intentan imitar las reacciones humanas, sin embargo, IBM comenta que *Watson Assistant* lo puede hacer cuando busca una respuesta en una base de datos, cuando pide más claridad en la pregunta y cuando pide el apoyo de un ser humano. *Watson Assistant* combina aprendizaje automático, comprensión del lenguaje natural, y herramientas de diálogo integradas para crear la conversación que fluye entre las aplicaciones y los usuarios. Se puede usar para crear interfaces de lenguaje natural para sus aplicaciones, construir agentes artificialmente inteligentes que se comunican con los usuarios a través del lenguaje natural, haciendo las aplicaciones más accesibles y de fácil uso para mantener una conversación con seres humanos. No se requiere codificación para configurar un *bot*. Cuenta con ayuda en línea integrada y se puede crear un *bot* en minutos. *Watson Assistant* requiere muy pocos datos para entrenamiento de una intención, con tan solo cinco ejemplos, el asistente de Watson puede comprender la intención, navegar desde un catálogo de servicio al cliente ya configurado y paquetes de contenido de la industria, para ahorrar tiempo. *Watson Assistant* tiene un paquete de diálogo básico que sirve en cualquier industria y existen paquetes de contenidos específicos para áreas de servicio al cliente en general, banca, tiendas minoristas, energía, servicios públicos, seguros y telecomunicaciones. Estos paquetes de contenido permiten concentrarse en

entrenar a su asistente en los detalles del negocio en lugar de tener que preocuparse por el conocimiento general.³⁴¹

- b. *Watson Discovery* (Descubrimiento): Busca dar respuestas, supervisar tendencias y patrones de servicio con el motor de información de la nube híbrida de IBM. *Watson Discovery Service* es una API que brinda la capacidad de recolectar, procesar, ordenar, enriquecer, indexar, buscar, almacenar y explorar sus datos, tanto estructurados como no estructurados. Primero, Watson recolecta y estandariza todos sus datos, sin importar el tipo. Posteriormente, los enriquece utilizando capacidades de comprensión del lenguaje natural. Analiza sentimientos y emociones, extrae entidades con nombre y hace etiquetado de conceptos. Crea nuevas formas de segmentación de datos. Se puede extraer, en segundos, sentimientos de los clientes con base en el historial de chat. No solo se puede hacer procesamiento de lenguaje natural. Al hacerlo puede trazar series temporales, realizar agregaciones, correlacionar variables. Puede recolectar y enriquecer rápidamente datos para construir aplicaciones cognitivas basadas en la nube híbrida, que conectan datos estructurados y no estructurados. *Watson Discovery Service* está disponible en *IBM Cloud*. Puede aprovechar los SDK³⁴² de Watson, las herramientas basadas en la web y las API de servicio de Watson Discovery para aportar información cognitiva a sus aplicaciones personalizadas. Permite usar colecciones preconstruidas, como *Watson Discovery News*. Permite investigar datos en un formato de lenguaje natural o en un lenguaje de consulta que buscará condiciones booleanas y frases. Una vez que se encuentran los datos, Watson Discovery utiliza algoritmos de aprendizaje automático para ordenar las respuestas de modo que las respuestas más útiles se enumeren primero. Discovery ofrece una ventaja inicial en análisis al proporcionar líneas de

³⁴¹ Para obtener más información sobre Watson Assistant, visite: [Asistente de Watson](#)

³⁴² SDK es el acrónimo de “Software Development Kit” (Kit de desarrollo de software). El SDK reúne un grupo de herramientas que permiten la programación de aplicaciones móviles. Este conjunto de herramientas se puede dividir en 3 categorías: (1) SDK para entornos de programación o sistemas operativos (iOS, Android, etc.); (2) SDK para el mantenimiento de aplicaciones; y (3) SDK de marketing y publicidad.

tiempo, gráficos y agregaciones. Todo respetando la privacidad ya que cuando se utiliza el servicio de descubrimiento de Watson, se mantiene el control y la propiedad de los datos, y el usuario decide qué datos utilizar y si desea que se hagan públicos. Adicionalmente, la explotación de datos sobre redes sociales, blogs, informes de investigación y correo electrónico empresarial se puede vincular con la función de comprensión del lenguaje natural de Watson que ofrece un conjunto de características para el análisis de textos (conocida como *Watson Natural Language Understanding*). IBM comenta que el servicio de comprensión del lenguaje natural de Watson permite entender los temas mencionados en el texto analizando palabras clave, conceptos y categorías. Puede extraer nombres de personas, lugares y organizaciones mencionados en el texto. Los usuarios también pueden entender el sentimiento que expresan las personas incluyendo sentimientos hacia una palabra clave, entidad o frase específica; además, pueden llevar el análisis de sentimientos un nivel más profundo y tomar emociones distintas detrás de lo que la gente escribe, al saber si el texto es positivo o negativo. La comprensión del lenguaje natural de Watson se vuelve aún más potente cuando se combina con otros servicios de *Watson Developer Cloud*. Por ejemplo, el usuario puede agregar imágenes y videos para su análisis o usar los servicios de voz para extraer información de los datos transcritos de los archivos de audio. Los datos son siempre mantenidos y controlados por su propietario.³⁴³

- c. *Watson Studio (Data Science)*: Permite crear, entrenar, implementar y administrar modelos de IA, preparar y analizar datos en un solo entorno integrado. Sin embargo, IBM entendió que extraer el valor total de los datos no es un proceso fácil. El primer reto está en qué hacer con los datos que se tienen y cómo limpiarlos, analizarlos y generar visualizaciones para, posteriormente, utilizarlos en la construcción e implementación de modelos de aprendizaje automático. Watson Studio ofrece una ciencia de datos

³⁴³ Para conocer más sobre Watson Discovery, se puede consultar: [Watson Discovery](#) [15 de julio de 2022]

colaborativa, fácil de usar y con un entorno de aprendizaje automático para construir y entrenar modelos, preparar y analizar datos y compartir ideas. Watson es compatible para trabajar con herramientas de código abierto. El entorno único de Watson Studio también crea un flujo de trabajo que es increíblemente eficiente. Para que los científicos de datos puedan compartir activos y trabajar para resolver problemas dentro del sistema en lugar de comenzar desde cero cada vez que surge un nuevo problema. Los desarrolladores, por su parte, pueden usar esa eficiencia para el desarrollo de algoritmos de aprendizaje automático. En el área de aprendizaje profundo, Watson Studio admite algunos de los marcos más populares y pueden implementar ese aprendizaje profundo para ayudar a acelerar el modelado mediante su uso fácil. El modelador de redes neuronales integrado también ayuda a construir modelos con una interfaz gráfica simplificada incluso si no tiene los recursos dedicados para construir un modelo desde cero. Watson Studio puede ayudarlo a comenzar con las plantillas de modelado para áreas como reconocimiento visual, clasificación de idiomas y otras herramientas de los servicios de IBM. Watson Studio está integrado con el Catálogo de conocimiento de IBM Watson y la herramienta inteligente de descubrimiento, puede transformar datos y modelos en recursos empresariales confiables, y colaborar sin comprometer el cumplimiento de las normas y la seguridad o los controles de acceso. El usuario puede entrenar a Watson Studio con reconocimiento visual. Puede personalizar sus modelos e implementarlos como API o Core ML utilizando herramientas de código abierto como Jupyter, Notebooks, Anaconda y RStudio. Watson Studio admite las bibliotecas de códigos más populares, así como con modelado visual de código con Neural Network Modeler para diseñar arquitecturas neuronales. Watson Studio puede descubrir, limpiar y transformar los datos usando Data Refinery, incluye una interfaz intuitiva de arrastrar y soltar que permite a un no programador acelerar el proceso de construcción del modelo seleccionando visualmente, configurar, diseñar y autocodificar redes neuronales, desde desarrollo y capacitación hasta producción y evaluación. Watson Studio

rastrea sus modelos a lo largo del tiempo para asegurarse de que tenga el mejor rendimiento para cualquier tarea, utilizando las mejores soluciones en todo el ciclo de vida de sus modelos de aprendizaje automático.³⁴⁴

- d. *Watson Machine Learning (Data Science)*: Permite usar sus propios datos para crear, entrenar e implementar modelos de aprendizaje automático y aprendizaje profundo. *Watson Machine Learning* brinda a las computadoras la capacidad de aprender sin programación. El servicio de aprendizaje automático de Watson, afirma IBM, es un conjunto de *API REST* a las que puede llamar cualquier lenguaje de programación para desarrollar aplicaciones que tomen decisiones más inteligentes, resolver problemas difíciles y mejorar los resultados del usuario. *Watson Machine Learning* está integrado con *Watson Studio*, permitiéndole integrar AI en sus aplicaciones y procesos. Puede crear diferentes modelos, usar paneles para comparar resultados y ejecutar experimentos automatizados y modelos de autoaprendizaje con el motor integrado *Watson Machine Learning*. *Watson Machine Learning*, puede crear modelos de aprendizaje automático utilizando herramientas de modelado visual e identificando rápidamente patrones; obtiene información para la toma de decisiones más rápidas. Puede utilizar los programas Jupyter, Python, R y Scala, o trabajar con R, en R Studio, para construir modelos analíticos sofisticados entrenados con sus propios datos. Puede conectarse a más de 300 diferentes almacenes de datos de terceros y aprovechar el poder de los datos compartidos para el desarrollo colaborativo. Puede diseñar redes neuronales complejas, luego experimentar e implementar modelos de aprendizaje profundo optimizados. Es fácil de escalar, independientemente del entorno en el que se encuentre trabajando en la nube híbrida, privada o híbrida. Simplifica y automatiza gran parte del proceso de capacitación e implementación, haciendo que sus productos integrados de IA estén disponibles más rápidamente. Tiene una variedad de modelos previamente entrenados y varios conjuntos de datos

³⁴⁴ Para mayores detalles consúltese: [Watson Studio](#)[15 de julio de 2022]

abiertos, lo cual le da una retroalimentación continua y el monitoreo del rendimiento lo ayuda a administrar su ciclo de vida.³⁴⁵

- e. *Watson Visual Recognition* (Visión): permite etiquetar, clasificar y entrenar contenido visual de forma rápida y precisa mediante el aprendizaje automático. La vida del ser humano es interpretación y lo hacemos de manera natural y al instante mediante la información visual y con el lenguaje común que podemos compartir. IBM piensa que todos los negocios de hoy luchan por entender y organizar la información de forma rápida y precisa. La información visual es difícil de interpretar. IBM Watson Visual Recognition, comprende datos visuales que son reconocidos por Watson. Puede convertir pilas de imágenes en información organizada. El reconocimiento visual puede aprender cualquier tipo de datos visuales. Simplemente se sube el conjunto de imágenes de entrenamiento (como, por ejemplo, imágenes de ventanas de automóviles rotas o imágenes de un modelo de automóvil como un híbrido, e imágenes de otros autos que no son híbridos y Watson las analiza). El reconocimiento visual puede aprender las diferencias entre las dos categorías cuando se encuentra con una nueva imagen que no había visto antes y puede distinguir en qué categoría pertenece. Los modelos de reconocimiento visual de Watson también se pueden incorporar a dispositivos móviles Apple iOS utilizando el marco CoreML de Apple. La personalización flexible significa que el servicio puede ser entrenado con clasificadores personalizados y conectado a otros servicios todo el tiempo manteniendo la privacidad y seguridad. Si se tienen muchas imágenes para analizar en un corto período de tiempo, el servicio de reconocimiento visual puede ser escalado para manejar cargas de trabajo pesadas. El servicio de reconocimiento visual genera palabras clave que describen lo que ve en las imágenes, e incluye una serie de características tales como la capacidad de identificar miles de diferentes objetos comunes que son encontrados en cuadros de video o imágenes fijas. El servicio también tiene la capacidad de

³⁴⁵ Para mayor información consúltese: [*Watson Machine Learning*](#) [15 de julio de 2022]

detectar rostros humanos e identificar una multitud de cualidades como el género, y edad o rasgos faciales. Hay otros modelos disponibles. para objetos y categorías como comida, texto e imágenes explícitas. Para 2021, Visual Recognition fue modificado y en su lugar se puso en su versión demo Remote monitoring, no intervendrá en temas de seres humanos solo en supervisión o inspección visual de objetos.³⁴⁶

- f. *Watson Knowledge Catalog* (Descubrimiento): Impulsa la colaboración y transforma los datos. IBM comenta que las empresas no siempre pueden encontrar y acceder fácilmente a los datos que necesitan. Muy a menudo pasan más del 80 por ciento de su tiempo localizando los datos correctos y se queda solo con el 20 por ciento de su tiempo para explotarlos. Además, en muchas ocasiones la complejidad de los tipos de datos (nuevos y emergentes), obliga a un examen y clasificación más profundo. Todo, sin contar que existen muchas regulaciones sobre cómo puede usar y compartir datos que hacen más lento este proceso de uso de datos para el crecimiento del negocio. IBM construyó Watson Knowledge Catalog que es un catálogo colaborativo de activos empresariales que está diseñado para descubrir y

³⁴⁶ <https://www.ibm.com/products/maximo/remote-monitoring#demo> A partir del 8 de junio de 2020, IBM ha anunciado que suspende su servicio de reconocimiento facial. Si uno entra a Watson esa función quedó desactivada para el uso público. El director ejecutivo de IBM, Arvind Krishna, envió una carta al Congreso de los Estados Unidos en la que hace propuestas para promover la igualdad racial. También compartió que IBM ha puesto fin a sus productos de software de análisis y reconocimiento facial de uso general (derivado del caso George Floyd). Cfr. <https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/> Amazon y Microsoft hicieron algo similar, pueden consultarse: <https://www.aboutamazon.com/news/policy-news-views/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition> y <https://www.iproun.com/innovacion/14477-por-que-microsoft-no-vendera-su-software-de-reconocimiento-facial-a-la-policia> [15 de julio de 2022] A principios de abril de 2021 Facebook anunció el proyecto “Casual Conversations Dataset” en el que por primera vez se pide autorización a las personas para el uso de su imagen y voz y se les paga por ello. Algo que los especialistas han estado exigiendo; también abrió un espacio open source de IA. Para conocer más de estos temas consúltase: <https://ai.facebook.com/datasets/casual-conversations-dataset/> y https://ai.facebook.com/research?utm_source=hp#fundamental-and-applied [15 de julio de 2022] Una controvertida herramienta de reconocimiento facial diseñada para la vigilancia se ha implementado en los Estados Unidos con poca o nula supervisión. Se llama “Clearview AI” y sirve para buscar rostros de personas. Según reportes de la asociación BuzzFeed News se utilizó para identificar a los manifestantes de Black Lives Matter, insurrectos del Capitolio, delincuentes menores, entre otros. Un tema muy polémico sujeto a regulación. Consúltase: Surveillance Nation, BuzzFeed, News, 9 de abril de 2021 en: <https://www.buzzfeednews.com/article/ryanmac/clearview-ai-local-police-facial-recognition?s=03>

catalogar enormes volúmenes de datos complejos y activos analíticos repartidos por múltiples fuentes y proporcionan funciones avanzadas para trabajo colaborativo con datos que ayudan a impulsar un nuevo crecimiento empresarial basado en inteligencia artificial. Para ayudar a combatir el problema de trabajar con grandes cantidades de datos que se extienden por toda la organización, *Watson Knowledge Catalog* puede conectarse a un conjunto diverso de fuentes de datos, locales o en la nube híbrida, incluyendo *Alexnet*, NoSQL, sistemas de archivos Hadoop, entre otros, para que se tenga la información correcta donde quiera que se encuentre y de forma rápida. *Watson Knowledge Catalog* utiliza el poder de la tecnología para enriquecer el índice de metadatos de información de la organización con clasificación, perfil y contenido de evaluación de calidad y proporciona recomendaciones inteligentes sobre los activos de datos más relevantes. Los catálogos de conocimiento de Watson tienen capacidades activas de aplicación de políticas para el control de accesos a la información y soporte automático y dinámico para mantener la confidencialidad de la información.³⁴⁷

- g. *Watson Knowledge Studio* (Descubrimiento): Proporciona una manera de enseñar a Watson el lenguaje de la industria o el servicio específico, con modelos personalizados que identifican entidades y relaciones exclusivas de su industria. Knowledge Studio proporciona herramientas fáciles de utilizar para anotar documentación de dominio no estructurada, y utiliza estas anotaciones para crear un modelo de aprendizaje automático personalizado que comprenda el idioma del servicio o producto específico. IBM comenta que la exactitud del modelo va mejorando mediante pruebas repetitivas, que finalmente da lugar a un algoritmo que puede aprender de los patrones que ve y reconoce dichos patrones en grandes recopilaciones de documentos nuevos. Puede desplegar el modelo de aprendizaje automático finalizado en otras soluciones cognitivas basadas en la nube híbrida de Watson para buscar y extraer las menciones de relaciones, incluidas las correferencias de

³⁴⁷ Para mayor información puede consultarse: [Catálogo de conocimiento de Watson](#)

las entidades. Las tecnologías tradicionales se han basado únicamente en la programación o crear reglas para personalizar los sistemas de lenguaje natural. IBM *Watson Knowledge Studio*, aprende el área de interés específica a través de ejemplos y sin escribir ningún código. Lo que solía durar años, ahora dura semanas. Watson se puede usar por los expertos de cualquier campo para entrenarlo en temas específicos. Para acelerar el entrenamiento, puede realizar anotaciones previas, documentos que utilizan la comprensión del lenguaje natural de Watson, especificar reglas o importar diccionarios. Luego puede ajustar los resultados para cumplir con sus requisitos. *Knowledge Studio* crea un modelo de aprendizaje automático basado en el entrenamiento que el usuario le proporcione. Watson puede entonces aprender solo a identificar entidades y relaciones que van más allá de su entrenamiento. *Watson Knowledge Studio* tiene flujo de trabajo colaborativo incorporado, que permite múltiples asignaturas con expertos en la materia para trabajar juntos, contribuyendo a una mejor y más precisa y soluciones desarrolladas más rápidamente. No requiere experiencia en ciencia de datos por parte del usuario o ni tampoco de lenguajes de programación.³⁴⁸

- h. *Watson OpenScale* (Data Science): Proporciona información sobre el estado de la IA, recomienda los pasos para mejorar los resultados y organiza las tareas para remediar problemas relacionados con el rendimiento, la precisión y la equidad. Es una plataforma que ayuda a las empresas a monitorear y ajustar el rendimiento de la IA a lo largo de todo el ciclo de vida. Por ejemplo, una compañía de seguros que desea utilizarlo para procesar reclamos. IBM dice que una característica de hacerlos explicables es que les permite obtener respuestas detalladas en términos que cualquiera pueda entender por qué se rechaza. La trazabilidad les permite volver sobre el proceso paso a paso y volver hacia atrás los documentos y datos que extrajo la IA. Si el sesgo ha influido en un resultado, una característica de sesgo señala cómo ocurrió y lo mitiga automáticamente. *Watson OpenScale* automatiza aspectos clave del ciclo de vida de la IA. Esto permite que crear y escalar

³⁴⁸ Watson Knowledge Studio Comprensión del lenguaje natural de Watson

modelos de IA en evolución se haga mucho más simple. Dentro de Watson OpenScale, existe una tecnología revolucionaria llamada NeuNetS (abreviatura de síntesis de redes neuronales), puede construir modelos rápida y automáticamente a partir de sus últimos datos mientras recomienda nuevos modelos para ser aplicados en su negocio. Como Watson OpenScale está diseñado para ser abierto, se puede conectar fácilmente a nuevos modelos de IA sin importar cómo o dónde se construyó. IBM *Watson OpenScale* pone a las empresas el control completo del ciclo de vida de su IA, dándoles la confianza para escalarlo y hacer decisiones rápidas, con posibilidad de quitarle lo perjudicial.³⁴⁹

- i. *Watson Natural Language Understanding* (Descubrimiento): Ofrece un conjunto de características para el análisis de texto que extrae entidades, relaciones, palabras clave, roles semánticos de 13 idiomas. Es una colección de API que ofrecen análisis de texto a través del procesamiento del lenguaje natural. Además, puede crear un modelo personalizado para algunas API para obtener resultados específicos que se adapten a su dominio. IBM comenta que este sistema es solo para fines de demostración y no está destinado a procesar datos personales. No se deben ingresar datos personales en este sistema, ya que puede no tener los controles necesarios para cumplir con los requisitos del Reglamento General de Protección de Datos (UE) 2016/679. *Watson Natural Language Understanding*, permite a los desarrolladores analizar las características semánticas de entradas de texto, incluidas las categorías, los conceptos, la emoción, las entidades, las palabras clave, los metadatos, las relaciones, los roles semánticos y el sentimiento. Además, permite analizar texto de páginas web mediante programación. Cuando se proporciona HTML sin especificaciones o un URL de acceso público en la solicitud de API, el servicio intenta centrar el análisis en el contenido principal de la página, como por ejemplo en el texto de un artículo de noticias o de una publicación de un blog. De forma predeterminada, el servicio limpia las páginas web antes de que se analicen

³⁴⁹ Consúltense [Watson OpenScale](#) [15 de julio de 2022]

e intenta eliminar el contenido que generalmente no se desea, como anuncios y otros textos que pueden interferir con el análisis del contenido principal de la página.³⁵⁰

- j. *Watson Speech to Text* (Habla): La tecnología de voz a texto ha existido por mucho tiempo, pero Watson está cambiando este modelo. La mayoría de las herramientas se centran en la transcripción de mensajes cortos y términos de búsqueda de audio de alta calidad. Pero ¿qué pasa con otras fuentes de audio, como llamadas telefónicas, reuniones y transmisiones? IBM *Speech to Text* utiliza técnicas avanzadas de modelado estadístico, desarrollado durante décadas y afinado con las herramientas de la computación cognitiva para transcribir tanto audio de alta calidad como de baja calidad y de una amplia variedad de fuentes. Determina automáticamente el resultado más preciso para palabras y frases y ofrece porcentajes de confianza y otros metadatos. IBM comenta que los centros de llamadas podrían transcribir millones de minutos de audio grabado automáticamente permitiéndoles extraer la información en las llamadas para identificar problemas y dar más valor a sus clientes y agentes. La transcripción precisa del diálogo permite prestar atención activamente a la discusión, ya sea de una grabación de audio o video o usando esta tecnología podría revisarse el contenido de una biblioteca de grabaciones de una empresa y combinarse con *Watson Developer Cloud*. *Watson Speech to Text* es un servicio basado en API que es especializado para convertir voz humana en texto con un formato de datos especial. Los resultados incluyen no solo el texto traducido, sino también traducciones alternativas junto con un puntaje competente para cada una de esas traducciones. Si se necesita detectores de voz para comprender palabras y frases menos utilizadas, quizás los específicos de su industria, el servicio puede ser entrenado para reconocer términos específicos del dominio. IBM proporciona acceso a una serie de kits de desarrollo de software disponibles en GitHub por trabajar con el servicio de voz a texto. Dado que el servicio está alojado en IBM *Cloud* es escalable, permitiéndole

³⁵⁰ Consúltase [Natural Language Understanding](#) [15 de julio de 2022]

tener múltiples servicios trabajando en concierto para traducir números muy grandes de discurso en texto. El servicio que ofrece IBM es altamente personalizable. Se puede entrenar a través de la API para reconocer muchas palabras y frases que puede ser específico para su caso de uso. *Watson Speech to Text* admite varias formas de conectarse al servicio con cualquiera en vivo transmisiones o audio pregrabado. IBM afirma que lo más importante de la industria es que todos los datos que pasan a través de este servicio de voz a texto son resguardados por el cliente o usuario. *Watson Speech to Text* transcribe automáticamente audio de siete idiomas en tiempo real o de fuentes grabadas.³⁵¹

- k. *Text to Speech* (habla): El servicio de texto a voz de Watson permite que sus sistemas hablen igual que los humanos. Está diseñado para convertir texto escrito en sonido natural audio en una amplia variedad de idiomas, dialectos y voces. El servicio de texto a voz incluye un potente sistema de generación de voz en tiempo real. Es altamente personalizable y se puede controlar la cadencia, la emoción y el tono del discurso producido. Fue creado para admitir una amplia variedad de casos de uso, puede sintetizar muchos idiomas y numerosas voces, incluyendo diferentes géneros y edades. El servicio de texto a voz admite el alfabeto fonético internacional y la notación de representación fonética simbólica de IBM, para permitirle construir pronunciaciones muy específicas para los términos, frases y palabras que pueden aparecer en su caso de uso. El servicio *Text to Speech* proporciona una *API REST*, que permite enviar texto para procesamiento del habla según sea necesario. IBM Watson tiene también una interfaz llamada *WebSocket* para transmitir salida de audio del texto al servicio de voz en tiempo real. También hay intercambio de recursos de origen cruzado o Soporte CORS integrado en el servicio de texto a voz. Esto significa que el servicio es accesible directamente por cualquier servidor configurado o aplicación sin tener que pasar a través de servidores específicos o servidores proxy. Para trabajar directamente con el servicio de texto a voz, las suscripciones al

³⁵¹ Para mayor información consúltese: [Speech to Text](#) [15 de julio de 2022]

sistema Watson incluyen acceso a kits de desarrollo de software y código GitHub. Todas estas herramientas se pueden utilizar conjuntamente. Watson Text to Speech le permite construir conversaciones muy específicas para términos, frases y palabras que pueden aparecer en cualquier industria.³⁵²

- I. *Watson Language Translator* (lenguaje): Si bien los traductores humanos son de alta calidad, también son muy caros y en muchas ocasiones no puede traducir en tiempo real dentro de las aplicaciones de software. Las empresas necesitan una forma asequible, automatizada de traducir información y necesitan poder ajustar sus modelos de traducción para tener en cuenta el idioma específico de su negocio. IBM comenta que el servicio de traductor de idiomas de IBM Watson permite a los desarrolladores de aplicaciones procesar directamente traducciones utilizando una *API REST* estándar simple. Además, para textos con terminología y lenguaje muy específicos, el servicio faculta a los usuarios con tres niveles de personalización. La personalización forzada del glosario le permite cargar pares de términos coincidentes en el idioma de origen y destino. Estos pares anulan por completo el modelo de traducción original. La personalización de frases paralelas le permite cargar una serie de frases coincidentes en el idioma de origen y destino. Esto difiere de la personalización del glosario porque no lo hace completamente, sino que anula el modelo de traducción original. La personalización a nivel de corpus permite a los usuarios cargar un archivo de texto sin formato, que contiene un cuerpo de texto en el idioma de destino, eso está relacionado con lo que estás traduciendo. Los clientes de IBM lo utilizan para traducir grandes cantidades de documentos internos, traducción también de texto externo como páginas web, traducción de comunicaciones en tiempo real como correos electrónicos, chats en línea y otra comunicación multilingüe. Los proveedores de servicios de idiomas también usan el servicio como una primera ronda de traducción para luego ser refinado por un traductor profesional. El servicio le permite traducir de todo, desde correos electrónicos, publicaciones en redes sociales, materiales de marketing,

³⁵² Para mayor información consúltese [Texto a voz de Watson](#) [15 de julio de 2022]

búsquedas de patentes, e incluso la interfaz de usuario de su aplicación, cuenta con traducción automática neuronal, que admite la personalización en 14 pares de idiomas. También tiene el servicio de traducción de documentos de un idioma a otro mientras se preserva el formato de archivo y tipo. Puede elegir mejorar resultados futuros al permitir el servicio entrenar usando sus datos que no se comparte ni se hace público, o al optar por no participar los datos no serán utilizados para mejoras de servicio general. *Watson Language Translator* admite la personalización, que puede ser especialmente útil para traducir términos técnicos.³⁵³

- m. *Watson Natural Language Classifier* (lenguaje): Aplica técnicas de aprendizaje profundo. Hace predicciones sobre las mejores clases predefinidas para oraciones o frases cortas, como descripciones, reseñas en línea, descripciones de trabajo o habilidades clave enumeradas en currículums. Se puede utilizar el aprendizaje automático para analizar texto, etiquetar y organizar datos en categorías personalizadas; para analizar descripciones de productos y organizar productos en categorías en sitios web de compras en línea, para clasificar las vacantes de trabajo en grupos, o para analizar el lenguaje utilizado y clasificarlo por tipos. Se puede clasificar las redes sociales o los datos analíticos, ordenar volúmenes de contenido escrito, evaluar los resultados e identificar habilidades clave. Las funciones de herramientas lo ayudan a clasificar y analizar tus datos. Tiene la función de clasificación multipropósito que sirve para extraer texto, así como separar y agruparlo en varios niveles cambiando múltiples clasificadores. Se puede aprovechar el aprendizaje profundo de Watson como un servicio para entrenar un modelo de clasificación de lenguaje natural en sus datos a velocidades más rápidas y en mayores cantidades. IBM *Watson Natural Language Classifier* utiliza algoritmos de aprendizaje automático para devolver las principales clases predefinidas coincidentes para entradas de texto cortas. Puede ayudar a cualquier aplicación a entender el lenguaje de

³⁵³ Para conocer más sobre este tema consúltese: [Watson Language Translator](#) [15 de julio de 2022]

textos cortos y a hacer predicciones sobre cómo manejarlos. *Natural Language Classifier* puede hacer esto en muchas aplicaciones y sectores diferentes. Por ejemplo: en el sector Banca y finanzas, puede clasificar inversiones, riesgos y transacciones. En el sector educativo y gobierno: ordena texto o documentos en categorías. Es útil para entornos académicos, legales, organizaciones sin ánimo de lucro y otras organizaciones que requieren clasificación. En el sector comercio electrónico y venta al por menor: ayuda a los usuarios a elegir productos mediante la reducción de las opciones por tema. Permite etiquetar productos o identificar artículos fraudulentos. En el sector servicios: clasifica consultas de servicio, mensajes y respuestas para ayudar a resolver problemas y a desplegar soluciones más rápidamente. En las redes sociales: organizar tweets, correo electrónico, publicaciones y comparticiones en categorías o temas. En el sector laboral: ayuda al descubrimiento de talentos, analiza currículums y aplicaciones para obtener información detallada. *Natural Language Classifier*, admite inglés, árabe, francés, alemán, italiano, japonés, coreano, portugués (de Brasil) y español.³⁵⁴

- n. *Watson Compare and Comply*: Optimiza los flujos de trabajo de contratos para ahorrar tiempo y mejorar la precisión en su redacción, así como simplificar la supervisión de su cumplimiento. *Watson Compare and Comply* permite recopilar información sobre los documentos y las organizaciones que los generan. Se selecciona el contrato para procesar y examinar su contenido. Se compara cada elemento, cada oración dentro del documento y permite encontrar rápidamente datos de interés. Las categorías definen el tema del elemento, por ejemplo, se puede seleccionar tanto derechos como obligaciones. Luego puede moverse fácilmente entre estos elementos. En segundo lugar, busca el tipo de acción que a cada una de las partes le compete realizar. Por ejemplo, puede elegirse buscar todas las obligaciones que corresponden al comprador o todas las del vendedor. Se puede mover

³⁵⁴ Watson Natural Language Classifier utiliza el aprendizaje automático para analizar texto y etiqueta y organiza los datos en categorías personalizadas. Para obtener más información sobre los servicios de idiomas de Watson, visite: [Clasificador de lenguaje natural de Watson](#) [15 de julio de 2022]

hacia los elementos identificados y combinarlos para encontrar todas las obligaciones del comprador. Se puede convertir los documentos de formato PDF o a HTML. Watson Compare and Comply agiliza los flujos de trabajo de los contratos para ahorrar tiempo y mejorar la precisión, y simplificar los contratos. Identifica fechas, nombres, direcciones, condiciones, plazos y los asocia con otras categorías para que se puedan analizar a criterio del usuario.³⁵⁵

- o. *Computer Vision* (Visión): Utiliza la propagación hacia atrás con una red neuronal convolucional de aprendizaje automático. IBM fue pionero al construir en 1967 Mark I Perceptron y a partir de 1980 el uso amplio de redes neuronales con retropropagación. Watson tiene muchos modelos pre-entrenados. como, por ejemplo, modelos que pueden ayudarlo a reconocer diferentes tipos de alimentos o modelos que pueden ayudar a reconocer diferentes tipos de objetos. La base de datos *ImageNet* es la que se utiliza para esta función. IBM comenta que se pueden tomar fotos de una línea de alimentación y entrenar a Watson con el poder del aprendizaje de transferencia en las redes neuronales.³⁵⁶ A partir del 8 de junio de 2020, Amazon, IBM y Microsoft anunciaron que suspenderían sus servicios de visión y reconocimiento facial para uso público y de cuerpos de seguridad con motivo del caso George Floyd. El presidente de Microsoft, Brad Smith, dijo que la compañía no venderá software de reconocimiento facial a los departamentos de policía de Estados Unidos hasta que haya leyes vigentes que rijan el uso de dicha tecnología: "... hemos decidido que no venderemos reconocimiento facial a los departamentos de policía en los Estados Unidos hasta que tengamos una ley nacional basada en los Derechos Humanos que regirán esta tecnología."³⁵⁷ A un año de distancia, los congresistas

³⁵⁵ Para obtener más información sobre los servicios de cumplimiento de Watson, consúltase: [Watson compara y cumple](#) [15 de julio de 2022]

³⁵⁶ Para obtener más información sobre los servicios de cumplimiento de Watson, consúltase: [Watson Visual Recognition](#). Por el momento está suspendido el servicio. [15 de julio de 2022]

³⁵⁷ Cfr. AP-Bloomberg, Las 3 grandes tecnológicas que negarán software de reconocimiento facial a la policía tras caso Floyd, *El Financiero*, 11 de junio de 2020, en <https://www.elfinanciero.com.mx/tech/las-3-grandes-tecnologicas-que-negaran-software-de-reconocimiento-facial-a-policia-tras-muerte-de-floyd/> [15 de julio de 2022]

norteamericanos no se han puesto de acuerdo para aprobar la ley para regular el uso de esta tecnología, se espera que su uso sea restringido para fines particulares y sometido a rigurosos protocolos de uso para las fuerzas de seguridad pública y seguridad nacional.

- p. *Personality Insights* (Empatía): El servicio Watson permite a las organizaciones comprender características de personalidad, necesidades y valores de las personas a través de los textos escritos. Obtiene información de las redes sociales de un individuo, de los datos empresariales o cualquier dato digital. Puede adaptar la experiencia de un cliente en función de cómo escribe usando esa escritura para adaptar su experiencia a sus necesidades, deseos o personalidad particulares. IBM afirma que se puede desarrollar una personalidad detallada, un retrato basado en el modelo de las cinco grandes personalidades³⁵⁸ (científicamente validado para evaluar la personalidad humana). El Servicio de Perspectivas de la Personalidad permite la escucha social, donde puede obtener ambos el significado semántico, así como un análisis más profundo del texto de una persona. Puede entender las preferencias de consumo de un cliente y adaptar los mensajes de la empresa en función de su personalidad, sus necesidades y sus intereses. Una vez que se haya utilizado el servicio para calificar las necesidades de personalidad de los clientes, puede personalizarse la presentación de productos e interfaces de usuario y otras características a las preferencias particulares de ese cliente. El Servicio de Perspectiva de la Personalidad también admite varios idiomas diferentes. El servicio puede interactuar individualmente con los clientes y personalizar su experiencia con la empresa a medida que las comunicaciones con su personalidad, necesidades y valor particulares se vayan incrementando. Cada cliente puede recibir un conjunto personalizado

³⁵⁸ Las cinco personalidades son: Apertura a la experiencia (*inventivo / curioso* vs. *consistente / cauteloso*) O; Conciencia (*eficiente / organizado* vs. *extravagante / descuidado*) C; Extraversión (*sociable / enérgico* vs. *solitario / reservado*) E; Amabilidad (*amigable / compasivo* vs. *desafiante / insensible*) A; Neuroticismo (*susceptible / nervioso* vs. *resistente / seguro*) N. Cfr. <https://psicologiymente.com/personalidad/5-grandes-rasgos-de-personalidad> [15 de julio de 2022]

de mensajes y se puede analizar el comportamiento del cliente a escala. Puede crear perfiles de personalidad para grupos y poblaciones de clientes que le permiten orientar productos con mayor eficacia, servicios y mensajes hacia esas poblaciones distintas.³⁵⁹

- q. *Watson Tone Analyze*: La gente usa varios tonos en comunicación diaria como alegría, tristeza, ira y amabilidad. Tales tonos pueden afectar la efectividad de comunicación en diferentes contextos. El analizador de tonos IBM Watson permite organizaciones para entender las emociones que se transmiten en el texto. Tone Analyzer utiliza análisis lingüístico para detectar tres tipos de tonos de comunicaciones; emoción, social y lenguaje; esto es, emociones como la ira, asco, miedo, alegría y tristeza. Propensiones sociales como la apertura, conciencia, extroversión, amabilidad, y rango emocional y estilos de lenguaje, por ejemplo, analítico, confiado y tentativo. El analizador de tonos, afirma IBM, puede realizar escucha social para extraer el significado emocional y efectivo del texto. Puede usarse para monitorear las redes sociales para entender lo que dicen los clientes sobre una marca, y para determinar a quién dirigirse con mensajes específicos. Permite personalizar comunicaciones de marketing con clientes, basados en su estado emocional actual. Puede usarse para mejorar el servicio al cliente, permitiéndole monitorear las reacciones para brindar un servicio de mejor calidad. Puede usarse para escalar conversaciones de clientes cuando se vuelven ásperas o encontrar oportunidades para mejorar los scripts de servicio al cliente, estrategias de diálogo y viajes de clientes. Se puede integrar con una amplia variedad de chatbots. Se puede acceder al analizador de tonos a través de una *API REST*, lo que permite una integración muy fácil con muchos otros servicios, incluidos muchos de los servicios que ofrece Watson, como Watson Assistant.³⁶⁰

³⁵⁹ Consúltese también: [Perspectivas de la personalidad de Watson](#) [15 de julio de 2022]

³⁶⁰ Consúltese también: [Analizador de tono Watson](#) [15 de julio de 2022]

B) Casos y evaluación del sistema

Caso 1 Bradesco

Bradesco es uno de los bancos más grandes de Brasil (5,200 sucursales). Antes que se utilizara Watson, cuando los empleados de una sucursal tenían preguntas sobre productos o servicios, llamaban a una oficina central, pero a menudo había una larga fila de espera para obtener atención telefónica. Esto desesperaba a los clientes y se generaban comentarios negativos. El banco adquirió Watson y ahora está atendiendo 62 productos y responde 283,000 llamadas al mes con una tasa de precisión del 95% (solo el 5% requiere atención de un ser humano). En la mayoría de los casos, los tiempos de respuesta se han reducido de 10 minutos a segundos. Esto ayuda a los empleados a tener una comunicación más enriquecedora con los clientes, porque tienen tiempo para dedicarse a temas de mayor relevancia para el cliente.³⁶¹

Caso 2 Coca Cola Company Watson y Core ML

Desde 2014, Apple e IBM han estado trabajando como socios para iniciar una nueva era de empresa inteligente. Watson y el marco de aprendizaje automático de Apple, Core ML, para aplicaciones nativas de iOS. Watson Services para Core ML ofrece aplicaciones iOS que brindan a los desarrolladores acceso a grandes cantidades de datos, tanto en su dispositivo como a través de la nube híbrida. Esto significa que los usuarios pueden acceder a información y conocimientos profundos directamente en su iPhone o iPad, incluso cuando no está conectado a una red. *The Coca-Cola Company* tuvo la oportunidad de explotar el valor de los servicios de Watson y ML para trabajar prototipos que transforman las capacidades de su personal técnico. Las funcionalidades iniciales que se analizan son la identificación de problemas de reconocimiento visual, el diagnóstico cognitivo y la realidad aumentada. Se

³⁶¹ Cfr. Customer response in seconds, not minutes
https://www.ibm.com/watson/stories/bradesco?mhsrc=ibmsearch_a&mhq=Caso%201%20Bradesco
[15 de julio de 2022]

implementan técnicas de campo para dar servicio y reparar máquinas dispensadoras de bebidas en todo el mundo. Antes de Watson, el técnico debía ser capaz de diagnosticar y corregir una enorme variedad de problemas, confiando en su experiencia y conocimientos personales. Si el sistema no era conocido o la pieza no era detectada, el técnico tenía que retirarse del lugar y consultar el tema para poder regresar a continuar la reparación. Esto causaba gastos y cierto estrés en el técnico. Además, muchos sitios se encontraban en ubicaciones rurales sin conectividad de datos, lo que significa que no hay acceso a soporte en línea y, por lo tanto, se reduciría la capacidad para realizar reparaciones. En todos los casos, el técnico necesitaba pasar tiempo buscando en bases de datos informativas y en los manuales de productos. Con las capacidades de Watson y Core ML, la información relevante se pone directamente en manos de un técnico en el momento en que la necesita, lo que le permite resolver el problema rápidamente. Coca-Cola utilizó los Servicios de Watson para Core ML para crear una aplicación que aprovecha el reconocimiento visual y la realidad aumentada para identificar problemas de equipos, diagnosticar problemas y solucionar problemas. A través de la aplicación, el técnico puede usar su cámara de iPhone o iPad para diagnosticar problemas del sistema a través de una superposición virtual e instrucciones guiadas extraídas de la nube híbrida, con latencia cero, e incluso en áreas sin conectividad de red. El reconocimiento visual de Watson en el dispositivo ayuda al técnico a identificar sistemas antiguos o poco diferenciados, o partes desconocidas, y a identificar el problema de inmediato. Luego, *Watson Discovery Service* ayuda a identificar posibles soluciones para los sistemas específicos y el tipo de mal funcionamiento. Utilizando ARKit, un marco de iOS con recursos para ayudar a crear experiencias de realidad aumentada para iPhone y iPad, los desarrolladores pueden integrar aplicaciones con modelos de realidad aumentada que ayudan al técnico a resolver problemas complejos en casos poco familiares. A medida que el técnico está trabajando se capturan datos. Luego, esos datos se envían a la nube híbrida, una vez que el dispositivo vuelve a estar en la red, para que Watson pueda aprender de la interacción y compartir el aprendizaje disponible para otros técnicos casi en tiempo real. Usando el sistema de reparación guiada, el técnico está capacitado

para resolver el problema la primera vez, aumentando la productividad y elevando el servicio al cliente, todo sin necesidad de pedir asistencia o reprogramar la reparación.³⁶²

Caso 3 Compañía Nacional de Petróleo de Abu Dhabi (ADNOC)

En el negocio del petróleo, la práctica común era que las compañías utilizaran a sus geólogos expertos -limitados en número- para describir toda la clasificación de rocas, y finalmente identificar las ubicaciones correctas para que los pozos sean perforados. Había una pregunta muy importante, ¿puede la inteligencia artificial ser realmente aplicada a un dominio tan técnico como la geología? ADNOC e IBM desarrollaron un proceso automatizado para analizar y categorizar muestras de rocas, acelerando el desarrollo de modelos digitales geológicos de yacimientos mediante la introducción de IA. Aunque la perforación y el escaneo de muestras de roca deben seguir siendo procesos manuales, la clasificación de imágenes es ahora mucho más rápida y automatizada. Watson puede analizar 527 imágenes por segundo. Analizar todas las muestras tomadas de un solo reservorio ahora solo toma minutos, no meses.³⁶³

Caso 4 Watson en uso en KONE

Los ascensores, las escaleras mecánicas y las puertas automáticas aseguran el paso de decenas de miles de personas todos los días en todo el mundo. Mantener a las personas en movimiento de manera rápida y segura requiere una enorme cantidad de esfuerzo para planificar el mantenimiento de rutina y resolver problemas imprevistos rápidamente. KONE presta servicios a más de un millón de ascensores y escaleras mecánicas que mueven a mil millones de personas por día. KONE utiliza

³⁶² Para saber sobre este caso consúltese: [IBM Watson Services for Core ML | IBM](#) [15 de julio de 2022]

³⁶³ Compañía Nacional de Petróleo de Abu Dhabi (ADNOC) <https://www.ibm.com/case-studies/abu-dhabi-national-oil-company-adnoc> [15 de julio de 2022]

Watson para ayudar a monitorear la condición del equipo y predecir los requisitos de mantenimiento y actualización, para ayudar al cliente a administrar su equipo durante todo su ciclo de vida. Esto significa menos tiempo de inactividad por fallas e información más detallada para los equipos de mantenimiento.³⁶⁴

Caso 5 Woodside Energy

Woodside es la compañía independiente de petróleo y gas más grande de Australia. En una industria que requiere una precisión absoluta, es fundamental que los ingenieros de Woodside eviten las conjeturas. Para garantizar la precisión, dependen en gran medida del contexto histórico y la información de procedimiento. Desafortunadamente, cada vez que un experto con años de conocimiento y experiencia se retira, esa experiencia sale con ellos. Entonces, la pregunta que se hizo para Woodside fue: ¿cómo retener el conocimiento de los expertos senior y hacer posible que los empleados junior lo ubiquen, analicen y aprendan de él? Para hacer que el conocimiento existente esté disponible en toda la empresa y preservar décadas y décadas de sabiduría colectiva, recurrieron a IBM y Watson. Para acumular la mayor cantidad de conocimiento posible para su solución Watson, el equipo de IBM-Woodside contrató al personal de toda la organización e incluso incorporó los aportes de los jubilados para tomar en cuenta todo el conocimiento. Antes de Watson, los ingenieros de Woodside pasaban hasta el 80% de su tiempo tratando de descubrir posibles soluciones o riesgos, y solo el 20% de su tiempo en el trabajo de Ingeniería real; con Watson, el tiempo dedicado a la investigación se ha reducido en un 75%, porque Watson permite un fácil acceso a las decisiones y conocimientos adquiridos por los propios empleados de Woodside.³⁶⁵

³⁶⁴ A billion people a day. Millions of elevators. No room for downtime. KONE and Watson IoT analyze data in elevators and escalators around the world to keep people moving smoothly, safely, and efficiently. <https://www.ibm.com/watson/stories/ca-en/kone-with-watson.html> [15 de julio de 2022]

³⁶⁵ Woodside Energy en: <https://www.ibm.com/case-studies/woodside-energy-watson-cognitive> [15 de julio de 2022]

Caso 6 Watson Discovery Service (Ciberseguridad)

IBM utiliza la inteligencia artificial de Watson para agregar un motor de información sobre amenazas cibernéticas a su plataforma de ciberseguridad: QRadar Security Analytics Platform. La mayoría de las vulnerabilidades en el tema de ciberseguridad son causadas por la llamada ceguera cibernética, es decir, los equipos dedicados a esta actividad no tienen forma de leer y analizar manualmente los enormes volúmenes de datos estructurados y no estructurados que necesitan para evitar una afectación a su organización. IBM construyó un centro de seguridad cognitiva con *QRadar Advisor* y *Watson*, combinando las capacidades de *Watson Discovery Service* y *Watson Knowledge Studio*. Los clientes que utilizan *QRadar Advisor* ha reportado que reduce el tiempo que sus analistas dedican a analizar las amenazas de nivel 1 hasta en un 50%, lo que les permite priorizar las correcciones más rápidamente y tomar acciones críticas antes; con el razonamiento cognitivo aplicado a grandes volúmenes de datos en constante cambio, *Watson* puede recopilar rápidamente evidencia de vulnerabilidades e incluso descubrir otras amenazas relacionadas con la ciberseguridad. Esto incluye archivos maliciosos, direcciones IP sospechosas, entidades poco confiables y, lo más importante, brinda información de todo esto a los analistas para una toma de decisiones más segura y efectiva.³⁶⁶

Caso 7 Watson for Oncology, Watson Clinical Trial, Watson for Genomics

Watson for Oncology, ayuda a crear regímenes de tratamiento centrados en el paciente; puede ayudar a identificar información sobre opciones de tratamiento centradas en el paciente basadas en evidencia científica, con contenido dinámico a través de la revisión de pautas científicas, ensayos académicos, artículos y datos de pacientes. El sistema arroja evidencia actualizada y de alto nivel para hacer

³⁶⁶ Papadopoulos, Leticia. How Watson AI is helping companies stay ahead of hackers and cybersecurity attacks, IBM, 14 de Agosto de 2017, en <https://www.ibm.com/blogs/watson/2017/08/how-watson-ai-is-helping-companies-stay-ahead-of-cybersecurity-attacks/> [15 de julio de 2022]

medicina personalizada en poco tiempo. El emparejamiento de *Watson Clinical Trial* puede ayudar rápidamente a identificar ensayos clínicos para la consideración de los médicos tratantes. Los ensayos clínicos requieren tiempo significativo para estudiarse, evaluar resultados y determinar qué pacientes se ajustan a qué estudio. Puede ser difícil identificar los ensayos y los pacientes que son elegibles, pero Watson eliminar la necesidad de hacerlo manualmente, esto significa que se puede ayudar a identificar pacientes y presentar a los médicos tratantes opciones de prueba de mayor facilidad y eficiencia. Watson puede también resumir los registros médicos del paciente, comparar las características relevantes a los protocolos actuales para determinar rápidamente qué ensayos son mejores y para qué pacientes. Las prácticas actuales revisan a cada paciente todos los días y esto es un proceso lento, con Watson se puede ahorrar ese tiempo. Existe también *Watson for Genomics* que puede ayudar a escalar el poder de la medicina al localizar información relevante para el perfil genético de una muestra con algún tumor; además puede omitir procesos manuales tardados de análisis de información que traen errores, sesgos o retrasos con información que integrar miles de artículos relevantes y proporcionar opciones terapéuticas centradas en el paciente. También puede obtener información para ayudar a los médicos a analizar los resultados de las pruebas genéticas y tomar decisiones sobre plan de tratamiento personalizado de cada paciente. *Watson for Genomics* ayuda a los proveedores a dar sentido a la enorme cantidad de información molecular que está disponible sobre sus pacientes respecto a tumores que los ayuda a crear las decisiones de tratamiento que deben seguir para el paciente en cada momento de atención.³⁶⁷

Watson incorporado en la nube híbrida de IBM cuenta con todo lo necesario para realizar cualquier proyecto de IA. Por ejemplo, para Knowledge Studio se cuenta con un editor de reglas³⁶⁸ con la opción de “Añadir diccionario”³⁶⁹, las palabras que se encuentren en el diccionario y coincidan con las entradas del

³⁶⁷ Supporting cancer research and treatment <https://www.ibm.com/watson-health/solutions/cancer-research-treatment> [15 de julio de 2022]

³⁶⁸ Ver Anexo 1

³⁶⁹ Ver Anexo 2

diccionario se anotan automáticamente con la clase asignada. Se puede también añadir una “expresión regular”³⁷⁰, es decir, una secuencia de caracteres que definen un patrón de búsqueda y un “glosario de términos y definiciones”³⁷¹, un apartado de técnicas para la resolución de problemas y provee en cada aplicación de un guía de funcionamiento que va describiendo producto

En marzo de 2021, Gartner³⁷², publicó un informe en el que evaluó a 20 proveedores de plataformas de ciencias de datos y aprendizaje automático (*Data Science and Machine Learning Platforms*, DSML) con el fin de brindar elementos técnicos para la toma de decisiones correctas de los ejecutivos de las empresas. Gartner define a los DSML como aquel producto principal que concentra una cartera de productos secundarios, componentes tecnológicos, bibliotecas y marcos de trabajo integrados coherentemente (incluidos los de código abierto, de fuente abierta y patentados) y que tiene como usuarios principales a profesionales de la ciencia de datos, incluidos científicos de datos expertos, científicos de datos ciudadanos, ingenieros de datos, desarrolladores de aplicaciones y especialistas en aprendizaje automático (ML).³⁷³ El estudio evalúa a los DSML bajo los siguientes criterios:

a) Producto principal:

- Debe estar lo suficientemente bien integrado para proporcionar un aspecto y sensación coherentes.
- Debe dar al usuario una experiencia en la que todos los componentes sean razonablemente interoperables para respaldar y canalizar el análisis.

b) Productos secundarios:

³⁷⁰ Ver Anexo 1

³⁷¹ Ver Anexo 3

³⁷² Gartner es una empresa de consultoría fundada en 1979, líder en investigación y asesoramiento de prácticamente cualquier industria o servicio que busca proporcionar a los altos ejecutivos de las empresas la información, el asesoramiento y las herramientas comerciales indispensables para lograr cumplir sus prioridades. Puede consultarse: <https://www.gartner.com/en> [15 de julio de 2022]

³⁷³ Gartner, *Magic Quadrant for Data Science and Machine Learning Platforms*, publicado el 1 de marzo 2021, en el sitio: <https://www.gartner.com/doc/reprints?id=1-25DOZD29&ct=210304&st=sb> [15 de julio de 2022]

- Debe ofrecer una combinación de funcionalidad básica y avanzada esencial para crear soluciones DSML (principalmente modelos predictivos y prescriptivos).
- Debe apoyar la incorporación de estas soluciones en los procesos comerciales, la infraestructura circundante, los productos y las aplicaciones.
- Debe apoyar el consumo sostenible de información derivada de la plataforma y ofrece funcionalidad para cuantificar y rastrear el valor de los proyectos de ciencia de datos.
- Debe admitir profesionales de la ciencia de datos con diversas habilidades (científico de datos es un título de trabajo aplicado de manera inconsistente y una distinción profesional): la base de usuarios de una plataforma DSML a menudo está formada por profesionales con diversos antecedentes técnicos y comerciales.
- Debe admitir múltiples tareas a lo largo del ciclo de vida de la ciencia de datos, que incluyen:
 - Comprensión del contexto empresarial y del problema
 - Ingestión de datos
 - Preparación de datos
 - Exploración de datos
 - Ingeniería de características
 - Creación y formación de modelos
 - Prueba de modelo
 - Despliegue
 - Supervisión
 - Mantenimiento
 - Gobernanza de datos y modelos
 - Inteligencia artificial explicable (XAI)³⁷⁴
 - Seguimiento del valor empresarial
 - Colaboración

³⁷⁴ IBM y GOOGLE son los que más están trabajando con Inteligencia Artificial Explicable.

Resultados del estudio del “cuadrante mágico” de Gartner



Fuente: Gartner (marzo, 2021)
https://www.gartner.com/resources/467300/467320/Figure_1_Magic_Quadrant_for_Data_Science_and_Machine_Learning_Platforms.png?reprintKey=1-25DOZD29

Para efectos de nuestra investigación resumiremos algunos de los resultados que se generaron respecto de IBM, MathWorks y SAS, como los competidores más fuertes del mercado, con una fortaleza muy antigua y, por otro lado, Microsoft, Google y Amazon Web Services, como los más innovadores que están compitiendo por alcanzar la supremacía del mercado.

a) *Amazon Web Services*³⁷⁵

Para Gartner AWS está considerado como un visionario en este cuadrante. Su visión es que los equipos de ciencia de datos utilicen toda la gama de la cartera de AWS y potencia de ML, con Amazon Sage Maker en su núcleo. AWS ofrece: SageMaker Studio IDE (que incluye piloto automático, portátiles, monitor de modelos, experimentos y depurador), Amazon EMR (incluido S3), AWS Glue, Amazon SageMaker Neo, Amazon SageMaker Ground Truth, Amazon SageMaker Clarify, Amazon SageMaker Data Wrangler, Amazon SageMaker Pipelines, AWS CloudWatch, AWS CloudTrail y otros. AWS está diversificado geográficamente y su base de clientes abarca muchas industrias y funciones comerciales.

Fortalezas

- Los usuarios pueden aprovechar directamente los servicios de inteligencia artificial preempaquetados de AWS (como Amazon Lex, Polly y Transcribe). SageMaker también está integrado de forma nativa con las numerosas herramientas de análisis y datos en la nube de AWS. SageMaker proporciona un amplio soporte para una amplia gama de bibliotecas y marcos de software de código abierto (OSS) populares y especializados.
- Amazon SageMaker y su cartera de soporte ofrecen el mejor rendimiento y escalabilidad de su clase. La plataforma admite una selección significativa de opciones de hardware optimizadas para varios marcos de aprendizaje profundo y aprendizaje automático, y presenta un modelo de precios de pago por uso sin tarifas mínimas ni compromiso inicial, lo que fomenta la experimentación.
- Amazon SageMaker Ground Truth admite el etiquetado de datos de entrenamiento, y la IA aumentada de Amazon (Amazon A2I) ayuda a crear flujos de trabajo óptimos para la revisión humana de los modelos implementados. AWS conecta a los clientes con proveedores del mercado

³⁷⁵ Para mayor información de AWS consulte su sitio en: <https://aws.amazon.com/es/> [15 de julio de 2022]

de terceros y la fuerza laboral de Amazon Mechanical Turk (MTurk) para el etiquetado humano de datos.

Precauciones

- La plataforma es más popular entre los codificadores; pero no es tan intuitiva para los usuarios sin conocimientos técnicos, en comparación con las herramientas líderes para los científicos de datos ciudadanos.
- Hay una incorporación rápida de nuevos componentes y servicios de AWS. Sin embargo, estas nuevas capacidades no están tan probadas ni son tan sólidas como las capacidades de otros proveedores para la preparación de datos, interfaces de usuario, colaboración y coherencia.
- La mayoría de los clientes de Amazon SageMaker operan en entornos puramente de nube. Algunas capacidades dentro de la cartera de AWS cambian o se vuelven más complicadas en entornos híbridos, multinube o locales. Sin embargo, la compatibilidad con múltiples nubes está evolucionando y, en la actualidad, la mayoría de los clientes administran datos, modelos y cargas de trabajo de aprendizaje automático dentro de AWS.

b) Google³⁷⁶

Para Gartner, Google es un visionario en este cuadrante. Google Cloud AI Platform es su plataforma DSML principal. La plataforma tiene un conjunto ampliado de componentes que incluye Cloud Data Fusion, Cloud AutoML, BigQuery ML, AI Platform Notebooks y TensorFlow. Google lanzó su plataforma de inteligencia artificial unificada en el primer trimestre de 2021 (después de la fecha límite para la evaluación en este Cuadrante Mágico). Las características y servicios clave que se lanzarán con esta nueva plataforma incluyen tablas AutoML, XAI, pipelines de plataforma AI y otros servicios MLOps. Google está diversificado geográficamente y su base de clientes abarca muchas industrias y diversas funciones comerciales.

³⁷⁶ Para mayor información consulte su sitio en: <https://cloud.google.com/> [15 de julio de 2022]

La integridad de la visión de Google se ve impulsada por el liderazgo intelectual en la investigación del aprendizaje automático y la inteligencia artificial responsable, así como por la hoja de ruta para su plataforma de inteligencia artificial unificada. La coherencia y la curva de aprendizaje de la plataforma de Google son aspectos clave a monitorear en 2022.

Fortalezas

- Google ha asumido una clara posición de liderazgo intelectual en el área de explicabilidad y responsabilidad de la IA. Google comparte y produce sus aprendizajes sobre estos temas a través de prácticas de inteligencia artificial responsables, mejores prácticas de equidad, referencias técnicas y otros materiales.
- El liderazgo de Google en la investigación de inteligencia artificial incluye el trabajo destacado de Google Research, Google Brain y DeepMind, así como contribuciones significativas en curso a becas, proyectos de código abierto y comunidades: se destacan TensorFlow, Kubernetes / Kubeflow y Kaggle.
- Google ha realizado un esfuerzo significativo para reorganizar y rediseñar no solo su plataforma DSML, sino también la forma en que lanza software. La plataforma de inteligencia artificial unificada buscará abordar problemas pasados de coherencia, interoperabilidad y facilidad de uso. Google también ha introducido etapas simplificadas de Introducción de nuevos productos (NPI) para proporcionar más previsibilidad y transparencia sobre los plazos de lanzamiento.

Precauciones

- Google está desarrollando capacidades para los profesionales de la ciencia de datos a un ritmo rápido. Esto significa un período de transición y aprendizaje para el mercado en general y para los adoptantes de su plataforma de inteligencia artificial unificada en particular. Los nuevos estándares y plazos de lanzamiento de productos de Google que se pusieron a prueba desde el 2021.

- Aunque Google ha realizado mejoras en términos de accesibilidad y aumento, su plataforma presenta una curva de aprendizaje empinada y requiere experiencia técnica. Es posible que se necesiten herramientas complementarias para los científicos de datos ciudadanos y los desarrolladores nuevos en ML.
- La mayoría de los clientes de *Cloud AI Platform* operan en entornos puramente de nube. Algunas capacidades cambian y pueden volverse más complicadas en entornos híbridos, multinube o locales. La compatibilidad con múltiples nubes está evolucionando y, en la actualidad, la mayoría de los clientes administran datos, modelos y cargas de trabajo de aprendizaje automático en *Google Cloud*. Los nuevos servicios como *BigQuery Omni* para ver datos en las nubes son indicativos de los próximos pasos de Google en el campo de la multinube.

c) *Microsoft*³⁷⁷

Para Gartner, Microsoft es un visionario en el cuadrante mágico. El producto principal considerado en este Cuadrante Mágico es *Azure Machine Learning* (Azure ML). La cartera de productos de apoyo para *Azure ML* incluye *Azure Data Factory*, *Azure Data Catalog*, *Azure HDInsight*, *Azure Databricks*, *Azure DevOps*, *Power BI* y otros componentes. Microsoft está diversificado geográficamente y su base de clientes abarca muchas industrias y diversas funciones comerciales. Microsoft obtiene la puntuación de capacidad de ejecución más alta de los grandes proveedores de nube. Microsoft tiene una sólida combinación de visión y funcionalidad personalizada para todo el espectro de profesionales de la ciencia de datos que contribuyen a los equipos multifuncionales.

³⁷⁷ Para mayor información consulte su sitio en: <https://azure.microsoft.com/es-mx/> [15 de julio de 2022]

Fortalezas

- *Azure ML* y su cartera de soporte ofrecen capacidades sólidas para las necesidades de la ciencia de datos empresarial. Las capacidades de MLOps incluyen un registro de paquetes y modelos y soporte para la creación optimizada de canalizaciones de ML reproducibles. *Azure ML* viene con capacidades diferenciadas de seguridad y gobierno y, combinado con los servicios de administración de *Azure Cloud*, admite capacidades de administración de costos y cuotas de cómputo.
- La visión y oferta actual de Microsoft para la ciencia de datos multipersona es más sólida que la de sus competidores más cercanos. *Azure ML* proporciona DSML aumentado y un diseñador de arrastrar y soltar para los científicos de datos ciudadanos, y opciones de SDK y portátiles flexibles para los científicos de datos expertos. El conjunto de productos auxiliares de Microsoft proporciona un entorno sólido en el que los ingenieros de datos, los ingenieros y arquitectos de ML, los desarrolladores corporativos y otros pueden contribuir al flujo de trabajo de DSML.
- Apertura y asociaciones: Microsoft va más allá del soporte generalizado para OSS populares al invertir y contribuir a una serie de proyectos destacados (como Open Neural Network Exchange [ONNX], *InterpretML* y *MLflow*). El producto y la asociación de *Azure Databricks* han sido exitosos para ambos socios. *Azure* también será el proveedor de nube preferido para la plataforma SAS Viya, que se integrará con los servicios *Azure* y *Azure ML*.

Precauciones

- *Azure ML* se basa en una variedad de servicios y módulos de *Azure* y puede trabajar con datos de cualquier fuente. Los clientes de *Azure ML* suelen utilizar *Azure Data Factory* para la integración y la transformación, *Azure Data Catalog* para la gobernanza y cualquier sistema DevOps (a menudo *Azure DevOps* o GitHub) para la integración dentro de los servicios web y otros

servicios. El soporte de esta cartera requiere una gran experiencia técnica y comprensión del ecosistema de Azure.

- La mayoría de los clientes de *Azure ML* operan en entornos puramente en la nube. Algunas capacidades cambian o se vuelven más complicadas en entornos híbridos, multinube o locales. Sin embargo, el soporte multinube de Microsoft está evolucionando y la mayoría de sus clientes administran datos, modelos y cargas de trabajo de ML dentro de Azure.
- Microsoft ofrece un soporte sólido para los científicos de datos ciudadanos y probablemente un costo total de propiedad más bajo para las capacidades de aumento, pero aún tiene espacio para mejorar, en comparación con los proveedores que se centran únicamente en la ciencia de datos. Las organizaciones que buscan ampliar su base de talentos en ciencia de datos deben comprender cuánto aumento ofrece el diseñador visual en Azure ML, a diferencia de los SDK y Power BI.

d) *IBM*³⁷⁸

Para Gartner, IBM es líder en el cuadrante mágico. Su producto principal para esta evaluación es IBM Watson Studio en IBM Cloud Pak for Data, una plataforma modular, abierta y extensible para datos e IA que combina un amplio conjunto de capacidades descriptivas, de diagnóstico, predictivas y prescriptivas. IBM está diversificada geográficamente y su base de clientes abarca muchas industrias y diversas funciones comerciales. Renovar su oferta ha llevado varios años y la competencia seguirá siendo feroz para IBM. Aun así, IBM ahora ofrece una solución moderna y completa que se basa en sus raíces en SPSS, ILOG CPLEX *Optimization Studio* y productos anteriores, y que se beneficia de una serie de innovaciones de *IBM Research*. Estos reflejan una visión completa.

Fortalezas

³⁷⁸ Para mayor información consulte su sitio en: <https://www.ibm.com/mx-es/cloud/watson-studio> [15 de julio de 2022]

- IBM Watson Studio ofrece una interfaz de flujo de trabajo visual o lienzo gráfico, así como una selección de portátiles, lo que permite a los ingenieros de datos, científicos de datos expertos y científicos de datos ciudadanos trabajar juntos en el mismo proyecto. Las actividades de la canalización de ML, desde la adquisición de datos hasta las operaciones, son compatibles con AutoAI y la colaboración, incluido un catálogo para compartir y reutilizar (meta) datos y modelos.
- La estructura modular de la plataforma IBM Watson Studio contiene, o puede ampliarse, múltiples componentes para el aumento de decisiones o la automatización. Estos componentes incluyen varios marcos de aprendizaje automático y otros marcos de inteligencia artificial, funciones de optimización, análisis espacio-temporal y de gráficos, funciones de lenguaje natural y análisis de video / imagen / audio (en modo por lotes o de transmisión). Además, al incluir IBM Decision Optimization, la plataforma admite el modelado de decisiones y la gestión de decisiones o el procesamiento de reglas.
- IBM ofrece un amplio soporte para la explicabilidad, el sesgo, la equidad, la precisión y el control de la deriva, los datos sintéticos y la privacidad diferencial. Su plataforma también proporciona un sólido soporte de gobernanza (y gestión de riesgos opcional), con linaje, políticas y reglas en su catálogo, así como seguridad contra adversarios.

Precauciones

- IBM Watson Studio ofrece automatización y aumento de múltiples actividades en la canalización de ML, incluida la selección de datos, imputación, visualización, transformación de características y modelado. Sin embargo, algunos competidores también aumentan el análisis de series de tiempo, por ejemplo, utilizando redes neuronales recurrentes y modelos de memoria a corto plazo.
- Watson Studio mejorado, ha superado a sus competidores. Sin embargo, los líderes de datos y análisis aún pueden encontrar escépticos a sus expertos

en ML sobre la innovación de Watson Studio y la capacidad de IBM para mantenerse al día en un mercado dinámico y competitivo.

- Aunque la cohesión del Watson Studio modular en IBM Cloud Pak for Data ha mejorado, sigue habiendo confusión entre los clientes potenciales en cuanto a qué productos y licencias se necesitan y para qué configuraciones.

e) *MathWorks*³⁷⁹

Para Gartner, MathWorks es líder en este Cuadrante Mágico. Sus dos productos principales son MATLAB y Simulink, pero solo MATLAB cumplió con los criterios de inclusión para este Cuadrante Mágico. MathWorks está diversificado geográficamente. Sus clientes son principalmente organizaciones de Ingeniería y centradas en activos. MathWorks demuestra una visión clara y liderazgo intelectual en industrias centradas en activos. Sus innovaciones se aplican, a escala, para grandes casos de uso destinados a resolver problemas del mundo real. MathWorks es uno de los pocos proveedores en el mercado de DSML que puede manejar implementaciones de IoT grandes, distribuidas y en tiempo real con un entorno continuo desde el borde hasta la nube, y desde el desarrollo hasta la simulación y operacionalización y viceversa.

Fortalezas

- MATLAB se encuentra entre las plataformas DSML más avanzadas para desarrollar, integrar y desplegar conjuntos de técnicas de IA dentro de una única solución (un enfoque que Gartner llama IA compuesta). MathWorks combina estas técnicas en una infraestructura flexible que admite entornos ampliamente distribuidos, desde el borde hasta el centro de datos y la nube.
- MathWorks se beneficia de una profunda experiencia en el dominio, que integra en su plataforma DSML. Desde mantenimiento predictivo hasta análisis de flotas, análisis de procesos de fabricación y gestión de riesgos, la empresa maneja idiosincrasias específicas de dominio dentro de su

³⁷⁹ Para mayor información consulte su sitio en: <https://la.mathworks.com/> [15 de julio de 2022]

plataforma, mientras desarrolla tecnologías y cajas de herramientas específicas para aplicaciones.

- La seguridad suele ser fundamental en los dominios centrados en activos en los que MathWorks está activo; no tienen tolerancia para operaciones no confiables. Más allá de la interpretabilidad, MathWorks permite a los ingenieros interactuar con modelos a través de aplicaciones web o entornos de simulación.

Precauciones

- MATLAB sigue siendo el dominio exclusivo de los ingenieros y científicos iniciados en la ciencia de datos, que básicamente utilizan portátiles para desarrollar modelos. Para ampliar el atractivo de su poderosa plataforma para los científicos de datos ciudadanos y los especialistas en operaciones y negocios, MathWorks tendrá que modernizar su interfaz de usuario y proporcionar funciones de desarrollo visual.
- MathWorks se mantiene por detrás de muchos de sus competidores, especialmente aquellos en el cuadrante de Líderes, en lo que respecta a la interpretación de modelos y la gestión de la equidad. Incluso su audiencia centrada en activos pronto requerirá mejores capacidades, por lo que la empresa tendrá que comenzar a enfocarse en este tema.
- A pesar del progreso en 2020, MathWorks se mantiene por detrás de muchos de sus competidores cuando se trata de expandir sus funciones DSML aumentadas, particularmente para la Ingeniería de funciones y la optimización de la implementación.

f) SAS³⁸⁰

Para Gartner, SAS es líder en este Cuadrante Mágico. SAS Visual Data Mining and *Machine Learning* (VDMML) es el producto principal evaluado para este Magic Quadrant. Como parte de la cartera de SAS Viya, VDMML se incluye en varios paquetes de productos en SAS Viya, a saber, SAS Visual *Machine Learning*, SAS Visual Data Science, SAS Data Science Programming y SAS Visual Data Decisioning. SAS está diversificada geográficamente y su base de clientes abarca muchas industrias y diversas funciones comerciales. SAS es el líder más antiguo en este Cuadrante Mágico. Mantiene una posición sólida y adaptable, dado su profundo conocimiento del mercado y su liderazgo intelectual en áreas clave como IA compuesta, MLOps e inteligencia de decisiones. La compañía anunció recientemente una asociación con Microsoft para apoyar una integración más cercana con Azure.

Fortalezas

- La larga trayectoria y experiencia de SAS en este mercado se ha ganado la confianza de los clientes. SAS ofrece soporte y capacidades de plataforma de nivel empresarial, junto con una visión sólida para las tendencias clave del mercado, incluida la inteligencia artificial compuesta, la inteligencia de decisiones y MLOps. La experiencia en el dominio incorporada en sus productos y servicios de consultoría permite a los clientes obtener valor de todo el ciclo de vida de la analítica.
- La última versión de SAS Viya ofrece un enfoque completamente nativo de la nube. Los clientes de SAS ahora pueden aprovechar todas las capacidades de Viya en una arquitectura flexible basada en contenedores que se ejecuta en la nube. SAS ofrece integraciones innatas con herramientas y lenguajes de código abierto populares para la gestión de datos, modelado y modelos.

³⁸⁰ Para mayor información consulte su sitio en: https://www.sas.com/es_mx/solutions/cloud/sas-cloud.html [15 de julio de 2022]

- SAS proporciona capacidades de modelado e Ingeniería de características automatizadas diferenciadas a través de la generación automatizada de tuberías. La experimentación se apoya a través de utilidades como el conjunto de acciones piloto de ciencia de datos y otros módulos. Para el ajuste automático de hiperparámetros, Model Composer utiliza una estrategia de búsqueda híbrida patentada.

Precauciones

- Los precios de SAS siguen siendo una preocupación para muchos clientes, quienes, por lo tanto, investigan alternativas menos costosas. El precio de VDMML históricamente ha sido fijado por el núcleo, pero un nuevo modelo de precios ha eliminado la restricción de capacidad del núcleo y el precio ahora se basa en el tipo de usuario. Los clientes de SAS deben trabajar con el proveedor para determinar si el nuevo modelo de precios es más adecuado para sus requisitos.
- Aunque SAS ha simplificado su cartera de productos, con más agrupaciones de productos aptas para el uso que reemplazan la selección a la carta, el conjunto completo de productos y complementos de SAS Viya sigue siendo complejo para que los usuarios naveguen. Sin embargo, para facilitar la navegación, SAS VDMML ahora forma parte de paquetes de productos que ofrecen interfaces de solo programación, así como paquetes que ofrecen interfaces de programación y visuales, de arrastrar y soltar.
- SAS necesita trabajar en la percepción de su cartera de productos. A pesar de una clara modernización, SAS todavía se percibe con frecuencia como un proveedor de software heredado y análisis avanzado tradicional. Las pequeñas y medianas empresas deben explorar estudios de casos de clientes en segmentos similares para comprender el uso actual de los productos SAS.

De acuerdo con Gartner, en este mercado de Google y Amazon ahora se siente a Microsoft como el competidor que busca la supremacía en términos de capacidades DSML en la nube. Sin embargo, las empresas grandes más antiguas de este sector, IBM, MathWorks y SAS, se mantienen firmes en sus áreas de éxito establecidas e innovan con ofertas modernas y estrategias adaptativas. El selecto grupo que aparece en este Cuadrante Mágico ha establecido sólidas bases de clientes, desempeño financiero y tecnología. La actividad de fusiones y adquisiciones en este mercado ha sido regular pero moderada. Es probable que los proveedores de este Cuadrante Mágico sigan adquiriendo empresas interesantes para completar sus plataformas, y siempre es posible realizar adquisiciones importantes y transformadoras. Sin embargo, a la fecha IBM Services diseña, construye y ejecuta los sistemas y servicios fundamentales que son la columna vertebral de la economía mundial.

Los clientes a los que IBM da soporte incluyen³⁸¹:

- 4 de las 5 principales aerolíneas por ingresos
- 8 de cada 10 operadores móviles líderes
- 8 de cada 10 mayores fabricantes de automóviles
- 4 de los 10 minoristas más grandes del mundo
- 7 de cada 10 compañías de seguros más grandes

La siguiente imagen muestra la infraestructura de IBM en el mundo (presente y futuro):

³⁸¹ BNP Paribas Signs an Agreement with IBM Services to further deploy its Cloud Strategy, Armonk, N.York, 22 de enero de 2019, en: <https://newsroom.ibm.com/2019-01-22-BNP-Paribas-Signs-an-Agreement-with-IBM-Services-to-further-deploy-its-Cloud-Strategy> [15 de julio de 2022]

Infraestructura para la nube híbrida (abril 2021)



Fuente: Educathon 2021

IBM le está apostando a la nube y ha creado tres modelos principales: Nube pública, nube privada y nube híbrida. Esos tres tipos de nube presentan características diferentes que ofrecen beneficios para cada tipo de aplicación en los negocios. Los expertos de IBM lo describen de la siguiente manera:

Nube Pública: El nombre “Pública” proviene de la idea que los recursos de un único servidor pueden ser usados por más de un usuario, y aunque el nombre pueda sugerir una idea contraria, existe un buen nivel de seguridad. Una de las maneras más conocidas de utilización de la nube pública es el “Pay as you go”, que se presenta también como un beneficio, dado que el usuario solo paga por la cantidad de recursos que usa. Por el lado empresarial, el beneficio comentado anteriormente es visto como una oportunidad de tener la posibilidad de usar esa tecnología con toda su flexibilidad y sin la necesidad de invertir en todos los costos que están involucrados en la creación de un data center los cuales son costo de Hardware, instalaciones, costos de la localidad, recursos energéticos, técnicos especialistas, entre otros. En el modelo de nube pública todos esos costos son de responsabilidad exclusiva del proveedor, que entrega un servicio listo para su uso, y ofrece también la escalabilidad, que permite la utilización de más recursos conforme la necesidad, sin un límite definido.

Nube Privada: Es un modelo en donde los servidores están disponibles exclusivamente al usuario final de dos maneras: *on-premises*, es decir, en donde la instalación de toda la infraestructura es hecha en la localidad del cliente y *off-premises*, en donde la infraestructura es instalada en la localidad del proveedor y son rentados exclusivamente al cliente. La nube privada requiere un nivel superior de responsabilidades por parte de la empresa contratante, adquiriendo la necesidad que ella posea profesionales capacitados para la gestión de todo el ambiente. El uso exclusivo por parte de una empresa ofrece un nivel superior de seguridad de entrada y permite que ella tenga todo el control de la infraestructura dedicada y la posibilidad de destinar los recursos de la mejor manera de acuerdo con su estrategia.

Nube Híbrida: Es una combinación estratégica de las nubes pública y privada. Esa combinación puede darse de diferentes maneras, sea el uso de una nube pública y una privada, dos nubes privadas, dos nubes públicas u otras asociaciones que sean definidas como ideal para cada empresa y su uso, siempre tomando en cuenta la necesidad de integrar dos o más ambientes de cloud. La creación de una arquitectura de nube híbrida puede ser pensada para solucionar tres problemas principales: La correcta conexión entre los ambientes de nube. La modernización de aplicaciones que van a ser usados en esos ambientes, con la intención de proporcionar el correcto funcionamiento. La seguridad, con la intención de ofrecer la mejor protección del ambiente como un todo incluyendo todo lo que estará contenido en ello.³⁸²

Como vemos, IBM Cloud es una **nube híbrida** con capacidades avanzadas de inteligencia artificial (*Machine Learning*) y procesamiento de datos. Las soluciones de nube híbrida de IBM Cloud ofrecen flexibilidad y portabilidad para crear una plataforma común que conecta los recursos locales de las organizaciones con todas las herramientas, software y redes que IBM ofrece a las empresas. IBM busca modernizar las aplicaciones existentes, crear y escalar las aplicaciones de la nube a petición de cada usuario, migrar las cargas de trabajo locales a la nube; acelerar la entrega de software y servicios, integrar aplicaciones y datos con otras nubes y acercar a Watson a todas las personas.³⁸³

³⁸² Cfr. Vennam, Sai, *Types of cloud computing*, IBM Cloud Learn Hub, 18 de agosto de 2020, en <https://www.ibm.com/cloud/learn/cloud-computing> [La traducción es propia] [15 de julio de 2022]

³⁸³ La contratación de servicios de la nube requiere de un soporte legal y el cumplimiento de un determinado Acuerdo de Nivel de Servicio (*Service Level Agreement*, SLA), que para este caso debe describir: (1) Las responsabilidades de cada una de las partes; (2) El nivel del servicio ofrecido (tasas de error permitidas, disponibilidad horaria, tiempos de respuesta y resolución o canales de comunicación); y, (3) Los procedimientos para el seguimiento y control del servicio y las sanciones por incumplimiento. Cfr. INCIBE, *Acuerdo pactado, contrato firmado. ¡Protege tu empresa!*, Instituto Nacional de Ciberseguridad, 15 de diciembre de 2020, Leon, en <https://www.incibe.es/protege-tu-empresa/blog/acuerdo-pactado-contrato-firmado-protege-tu-empresa> [15 de julio de 2022]

IBM está consciente que el futuro está en el manejo y producción de datos y que su utilización se requerirá por millones de personas y sistemas de IA. Para los fines de esta investigación, tomaremos como base el sector de los servicios de salud, en el que IBM ha invertido muchos recursos para ofrecer herramientas que permitan a los médicos mejorar la toma de decisiones en diagnósticos y tratamientos.

2) Caso Watson Health

Antes de entrar al caso *Watson Health*, consideramos de la mayor relevancia dimensionar el origen histórico de la medicina, cómo van surgiendo los principales avances tecnológicos y cómo surge la llamada *Healthcare technology*. Este seguimiento histórico comienza desde que el ser humano decide ser un animal sedentario y comienza a construir comunidades de personas establecidas permanentemente en un sitio. Existen evidencias de los primeros intentos documentados del oficio del médico y de las enfermedades en civilizaciones como Egipto y Mesopotamia alrededor del 1000 a. C., India alrededor del 750 a. C., Grecia del siglo VI a. C. y China alrededor del 100 a. C. ³⁸⁴

Al constituirse las civilizaciones se fueron documentando las enfermedades, sus remedios o curas y los costos (económicos y sociales). Enfermedades como la malaria, tuberculosis, cólera, viruela, sarampión, paperas, disentería fueron llegando al mismo tiempo que el crecimiento económico y la conquista de nuevos territorios; también las *epidemias* o pandemias como la peste bubónica y la gripe del COVID-19.

Se le atribuye al médico suizo Theodor Zwinger I (1533-1588)³⁸⁵, la búsqueda de diversos datos que permitieron entender la evolución de la medicina, desde la

³⁸⁴ Roy Porter, *The Cambridge Illustrated history of medicine*, Cambridge University Press, Cambridge, 1996, pp. 16-30.

³⁸⁵ Theodor Zwinger I (o el anciano) se le atribuye ser el autor de la primera enciclopedia denominada *Theatrum vitæ humanæ*, Basilea, 1565, contiene una amplia compilación de experiencias y

mitología griega hasta el siglo XVI. La medicina, se comenzaba a construir como algo más allá de la mera curación, como la posesión de un cuerpo específico de aprendizaje, teórico y práctico, que podría usarse para tratar a los enfermos.³⁸⁶

Aunque Theodor Zwinger I tenía razón al situar los orígenes de su tradición de la medicina en la antigua Grecia, los griegos no eran los únicos pueblos de la región del Mediterráneo oriental que podían afirmar haber inventado la medicina. La antigua Mesopotamia y Egipto tenían textos y tradiciones médicas que eran anteriores a los de Grecia. Las tradiciones de India, China y el Lejano Oriente, tienen el mismo derecho a la antigüedad.

El pensamiento médico de Galeno de Pérgamo (129-201)³⁸⁷, en particular, fue deificado desde el siglo II hasta toda la Edad Media hasta que llega el Renacimiento y la Revolución Científica. La medicina galénica tenía como debilidad lo anatómico, ante la creencia de que el cuerpo se debía respetar porque pertenecía a los dioses (se abstenían de cortar cuerpos muertos). La peste negra influyó en la decisión de la iglesia y los médicos de abrir los cuerpos para ubicar las causas de la muerte.³⁸⁸ Los defensores de la nueva ciencia, como René Descartes, denunciaron a Aristóteles por dotar falsamente a la naturaleza de vitalidad, inclinaciones, apetitos, voluntad, conciencia y propósitos (causas finales). En realidad, argumentaron, la naturaleza estaba formada por partículas o corpúsculos de materia inerte. Los movimientos de los planetas, o de las bolas que caen de la Torre de Pisa, no debían explicarse en términos de ningún deseo, sino a través de las leyes de la mecánica, de la materia en movimiento, que encontraron una expresión ideal en el lenguaje de

anécdotas históricas, sobre la imaginación, la memoria, las enfermedades mentales, la voluntad, las costumbres, las ciencias y las tecnologías de esa época.

³⁸⁶ Ibidem., p. 93 y 95

³⁸⁷ Fue un referente obligado para los estudiosos de la medicina europea a lo largo de más de mil años en temas como anatomía, fisiología, patología, farmacología y la neurología.

³⁸⁸ Fue el Papa Clemente VII quien aceptó la enseñanza de la anatomía por disección. Italia era el centro de la investigación científica.

la geometría y el número. La naturaleza era una máquina y los seres humanos también.³⁸⁹

Durante el siglo XVII, los humores se despreciaron y se descartaron. Gracias a las técnicas de disección otorgadas por el prestigioso anatomista flamenco Andreas Vesalius (1514-1564), la atención se centró de nuevo en las partes sólidas del cuerpo en su obra insuperable: *De humani corporis fabrica libri septem* (1543)³⁹⁰. El modelo de máquina tenía implicaciones profundas para las concepciones de la salud y la enfermedad: el bienestar se comparó con el funcionamiento de una máquina bien afinada y engrasada, y la enfermedad se describió como una avería mecánica, debido quizás a un bloqueo, escasez de combustible o fricción excesiva. El pionero de esta forma de pensar fue el filósofo y científico natural René Descartes (1596-1650) quien veía al cuerpo como una máquina. Descartes, nacido en Francia, pasó sus años creativos en la República Holandesa, desarrollando una Filosofía dualista que postulaba la distinción esencial entre mente y materia. Su enfoque naturalista desmitificó el cuerpo humano y lo dejó completamente abierto al análisis científico.³⁹¹

Para el siglo XVIII, la medicina y las investigaciones anatómicas continúan con los postulados de Descartes y Vesalius, desarrollándose jeringas, válvulas, tubos y demás dispositivos para disección que pudieran exponer el cuerpo en estructuras y fluidos. En 1710, se publica por Theodor Zwinger III (1658-1724) el *Theatrum praxeos medicae*³⁹². Una interesante compilación de diversos conceptos médicos, entre los que ya se encuentra la definición del cáncer: *CANCER, CARCINOS, CARCINOMA, eft tumor durus, plus minus dolens, lividus, & fubinde pluribus*

³⁸⁹ Roy Porter, *The Cambridge Illustrated history of medicine*, Cambridge University Press, Cambridge, 1996, pp. 93 y ss.

³⁹⁰ El contenido de la obra se divide en: Libro I: de los huesos y de las articulaciones. Libro II: de los músculos. Libro III: Corazón y vasos sanguíneos. Libro IV: Sistema Nervioso. Libro V: Órganos abdominales. Libro VI: Órganos de la región del tórax. Libro VII: cerebro. Cfr. <http://uvadoc.uva.es/handle/10324/15161> [15 de julio de 2022]

³⁹¹ Roy Porter, *The Cambridge Illustrated history of medicine*, Cambridge University Press, Cambridge, 1996, pp. 93 y ss.

³⁹² Cfr. Zwinger, Theodor, *Theatrum praxeos medicae*, Basilea, Johannis Brandmülleri, 1710, s/l, p. 176. Vea la imagen del “Anexo 5” de esta sección. Existe una versión pública que se puede consultar en: https://archive.org/details/b30550075_0001/page/n7/mode/2up [15 de julio de 2022]

varicofis venis praeditus» Vocatur etiam Lupus, quod vicinas fenfim carnes exedat, & Noli me tangere, quoniam acria medicamenta minime tolerat. Nafci poteft in omnibus ferè corporis noftri membris, in mammis praecipuè, labiis, oculis, collo, facie, pudendis, brachiis &c. Praecipuè autem glandulofa loca hujus diri Affectûs fubjectum evadere folent. El antiguo énfasis humoral se mantenía, pero con términos mecánicos e hidrostáticos.³⁹³

La llegada de la electricidad permitió a la medicina dar grandes pasos. La electrofisiología experimental y la aplicación de corriente eléctrica al cuerpo³⁹⁴ como medio para experimentar reacciones en nervios y músculos va generando nuevos conocimientos. Luigi Galvani, pionero en estos temas, publica *De Viribus Electricitatis in Motu Vusculari* (Sobre las potencias eléctricas en el movimiento de los músculos, 1792). En esa misma época Alessandro Volta, demostró que los músculos se contraen con estímulos eléctricos en su obra *Letters on Animal Electricity*, 1792.³⁹⁵ Ya para el siglo XIX el tema de la electricidad y su conexión con la vida y el movimiento del cuerpo llegan a inspirar a Mary Sheller para escribir *Frankenstein* (1818), cuya trama giraba en torno a la creación artificial de la vida a través de la experimentación fisicoquímica y los peligros de tal poder.³⁹⁶

Quizás sea Michel Foucault³⁹⁷, el autor que analiza con mayor precisión los orígenes de los sistemas de salud, la labor del médico y la incorporación del hospital como parte de la tecnología moderna.

³⁹³ Roy Porter, *The Cambridge Illustrated history of medicine*, Cambridge University Press, Cambridge, 1996, pp. 162.

³⁹⁴ El físico alemán Otto von Guericke (1602-1686) desarrolló la primera máquina electrostática para generar cargas eléctricas. El físico holandés, Pieter van Musschenbroek, en 1746 inventa el primer condensador eléctrico: la botella de Leyden. Un dispositivo eléctrico realizado con un envase de vidrio que permite almacenar cargas eléctricas, comportándose como un condensador o capacitor. La jarra se llena de agua, se tapa con un corcho y es atravesada en su centro por un cable con uno de sus extremos sumergido en el agua. Al conectarla a una fuente de energía estática la botella se cargaba y podía descargarse conectando su borne central a un punto de potencial cero.

³⁹⁵ Roy Porter, *The Cambridge Illustrated history of medicine*, Cambridge University Press, Cambridge, 1996, pp. 167.

³⁹⁶ *Idem*.

³⁹⁷ Michel Foucault, "Incorporación del hospital en la tecnología moderna" en *Educ Med Salud*, Vol. 12, no. 1, 1978, pp. 20-35.

Foucault se pregunta ¿Cuándo empezó a programarse el hospital como un instrumento terapéutico, instrumento de intervención en la enfermedad o el enfermo, instrumento capaz, por sí mismo o por alguno de sus efectos, de curar? Y nos da un panorama muy puntual del hospital como instrumento terapéutico que surge en Europa a finales del siglo XVIII. Alrededor de 1760, se crea una conciencia de que el hospital puede y debe ser un instrumento destinado a curar al enfermo y se busca a expertos para estudiar su estructura, sus actividades y los cambios que se deben hacer para ello.

John Howard y Jaques Rene Tenon son ubicados por Foucault como los primeros que estudian las características de los hospitales con el fin de construir un sistema de hospitales en Inglaterra y en Francia. Al hacer su recorrido por diversos hospitales de Europa, Howard y Tenon:

“daban cuenta del número de enfermos por hospital, de la relación entre el número de pacientes, del número de camas, del espacio útil de la institución, de la extensión y altura de las salas, de las unidades cúbicas de aire de que dispone cada enfermo, y de la tasa de mortalidad o de cura. También incluían una investigación de las relaciones entre fenómenos patológicos... Por ejemplo, Tenon investigaba en qué condiciones especiales se curaban mejor los casos hospitalizados por heridas y cuáles eran las circunstancias más peligrosas para ellos... Tenon estudiaba asimismo los recorridos, desplazamientos y movimientos en el seno del hospital, particularmente el trayecto que seguía la ropa blanca, sábanas, ropa vieja, trapos utilizados para curar a los heridos, etc. Investigaba quién transportaba este material y dónde se llevaba, se lavaba y se distribuía”.³⁹⁸

La aportación principal de Foucault, lo que hace especial su estudio, está en el recorrido histórico-descriptivo para detectar cómo surgen los hospitales, su evolución y el papel que fueron cobrando los médicos. Primero, nos deja claro que la imagen del hospital que hoy conocemos no corresponde con la imagen de hospital que funcionaba en Europa desde la Edad Media, porque su origen y función no iban enfocadas hacia la cura o la consulta médica ni tampoco a la recuperación de los enfermos. “La medicina era una profesión no hospitalaria.... el hospital era esencialmente una institución de asistencia para los pobres, pero al mismo tiempo era una institución de separación y exclusión. El pobre... era portador de

⁴⁰⁰ *Idem.*

enfermedades y posible propagador de estas... El hospital era un lugar para ir a morir... la institución servía para salvar el alma del pobre en el momento de la muerte... la salvación espiritual más que en el material, unida a la función de separar los individuos peligrosos para la salud general de la población.”³⁹⁹

En segundo lugar, la práctica médica, no necesitaba al hospital. La medicina medieval y hasta el siglo XVIII era individualista y solo comprendía el estudio de textos y recetas, pero no incluía una estancia práctica en el hospital. No había en la organización de los conocimientos hospitalarios la intervención del médico... En consecuencia, hasta mediados del siglo XVIII el hospital y la medicina siguieron siendo campos independientes”.⁴⁰⁰

El tercer punto que delimita Foucault está en el momento en que se medicalizó el hospital o en la llegada de la medicina al hospital. El factor principal -dice Foucault- lo encontramos en los hospitales marítimos y militares y el desorden económico que se generaba en su interior. En este espacio se traficaban mercancías, objetos preciosos, materias raras, especies, etc., procedentes de las colonias:

“Si los hospitales militares y marítimos se convirtieron en modelo... es porque con el mercantilismo las reglamentaciones económicas se hicieron más estrictas y también porque el precio del hombre aumentaba cada vez más... en esa época la formación del individuo, su capacidad, sus aptitudes, empiezan a tener un precio para la sociedad... el hospital militar se convierte en una cuestión técnica y militar importante... La reordenación de los hospitales marítimos y militares no partió de una técnica médica sino, esencialmente, de una tecnología que podría denominarse política: la disciplina... La disciplina es una técnica de poder que encierra una vigilancia perpetua y constante de los individuos... supone un registro continuo: anotaciones sobre el individuo y transferencia de la información en escala ascendente... La introducción de los mecanismos disciplinarios en el espacio confuso del hospital iba a permitir su medicalización. Las razones económicas, el precio atribuido al individuo, el deseo de evitar la propagación de las *epidemias* explica la fiscalización disciplinaria a que están sometidos los hospitales. Pero si esta disciplina adquiere carácter médico, si este poder disciplinario se confía al médico, se debe a una transformación del saber médico. La formación de una medicina hospitalaria hay que atribuirla, por un lado, a la introducción de la disciplina en el espacio hospitalario y, por el otro, a la transformación que en esa época experimenta la práctica de la medicina.” Desde el momento en que el hospital se concibe como un instrumento de cura y

³⁹⁹ *Idem.*

⁴⁰⁰ *Idem.*

la distribución del espacio se convierte en instrumento terapéutico, el médico asume la responsabilidad principal de la organización hospitalaria. A él se le consulta cómo se debe construir y organizar el hospital; por este motivo Tenon realizó la mencionada encuesta.⁴⁰¹

Los hospitales se convirtieron en lugares clave para la ciencia médica, comenzaron a producirse avances notables en la fisiología y la patología con los diversos laboratorios que existían. Los practicantes de la química orgánica, la microscopía y la fisiología del siglo XIX creyeron con razón que estaban dando a luz a una nueva empresa basada en el laboratorio. El hospital era un lugar para observar, el laboratorio para experimentar. La ciencia fue viento en popa en el siglo XIX, ganando mayores fondos públicos.

En 1865, Claude Bernard publica *Introduction a la Medecine experimental*, una exposición sistemática del método experimental para las ciencias biomédicas. La medicina hospitalaria tradicional, sostenía Bernard, tenía dos limitaciones clave. Como ciencia observacional, era puramente pasiva, similar a la historia natural. El progreso de la fisiología requirió la observación activa del experimentador en condiciones controladas. En el lecho del enfermo, había demasiados imponderables para permitir una comprensión precisa. Además, argumentó (en contradicción con Pierre Louis, Rene Laennec y su escuela), la lesión patológica en sí misma no era el origen sino el punto final de la enfermedad. El conocimiento fisiopatológico sólo podría cumplirse en el laboratorio, y solo mediante el uso de animales de laboratorio en ambientes controlados. No hay patología sin fisiología, insistió. La interacción de la fisiología, la patología y la farmacología constituían para él los cimientos de la medicina experimental, y cada una tenía que ser una ciencia de laboratorio.⁴⁰²

El siglo XX trajo desarrollos sin precedentes en biología, química y fisiología y la apertura de nuevas especialidades dentro de la ciencia médica. Sería imposible enumerar todos los principales avances de la ciencia médica, pero sí es posible

⁴⁰¹ *Idem.*

⁴⁰² Roy Porter, *The Cambridge Illustrated history of medicine*, Cambridge University Press, Cambridge, 1996, pp. 182.

esbozar algunos campos y avances destacados con el bienestar humano. La investigación microbiológica promovida por Louis Pasteur y Robert Koch, por ejemplo, condujo a la creación alrededor de 1900 de la inmunología.⁴⁰³

La palabra inmunidad, exención de una enfermedad en particular, se popularizó a medida que los investigadores se familiarizaban con las enigmáticas relaciones de infección y resistencia. Fascinado por las necesidades nutricionales de los microorganismos, Pasteur sugirió una dimensión nutricional a la resistencia de un huésped y la atenuación de un parásito: el microorganismo perdió su poder de infectar porque ya no podía florecer y reproducirse. Pasteur estaba más preocupado por la producción de vacunas que por las razones teóricas por las que las vacunas protegían (o inmunizaban). El examen microscópico de animales infectados con varios patógenos, incluido el bacilo del ántrax, mostró que los glóbulos blancos atacaban y parecían digerir la enfermedad.⁴⁰⁴

Visto el fenómeno de evolución de la ciencia médica en comparación con el enfoque de la esperanza de vida de las personas, encontramos que hubo un incremento considerable. Durante la Edad Media y todo el Renacimiento la esperanza de vida era de aproximadamente 30 años.⁴⁰⁵ Posteriormente, a principios del siglo XIX, la esperanza de vida comenzó a aumentar en los países industrializados, pero nunca pasó de 40 años. Las mejoras llegarían después de la Segunda Guerra Mundial. La década de los años 50 del siglo XX, muestra una esperanza de vida que superará los 60 años en Europa, América del Norte, Oceanía, Japón y partes de América del Sur. Pero en otros lugares, como África, un recién nacido alcanzaba la edad de 30 años. Existía una gran disparidad en este tema. Había países como Noruega que tenían una esperanza de vida de 72 años,

⁴⁰³ Ibidem., p. 189.

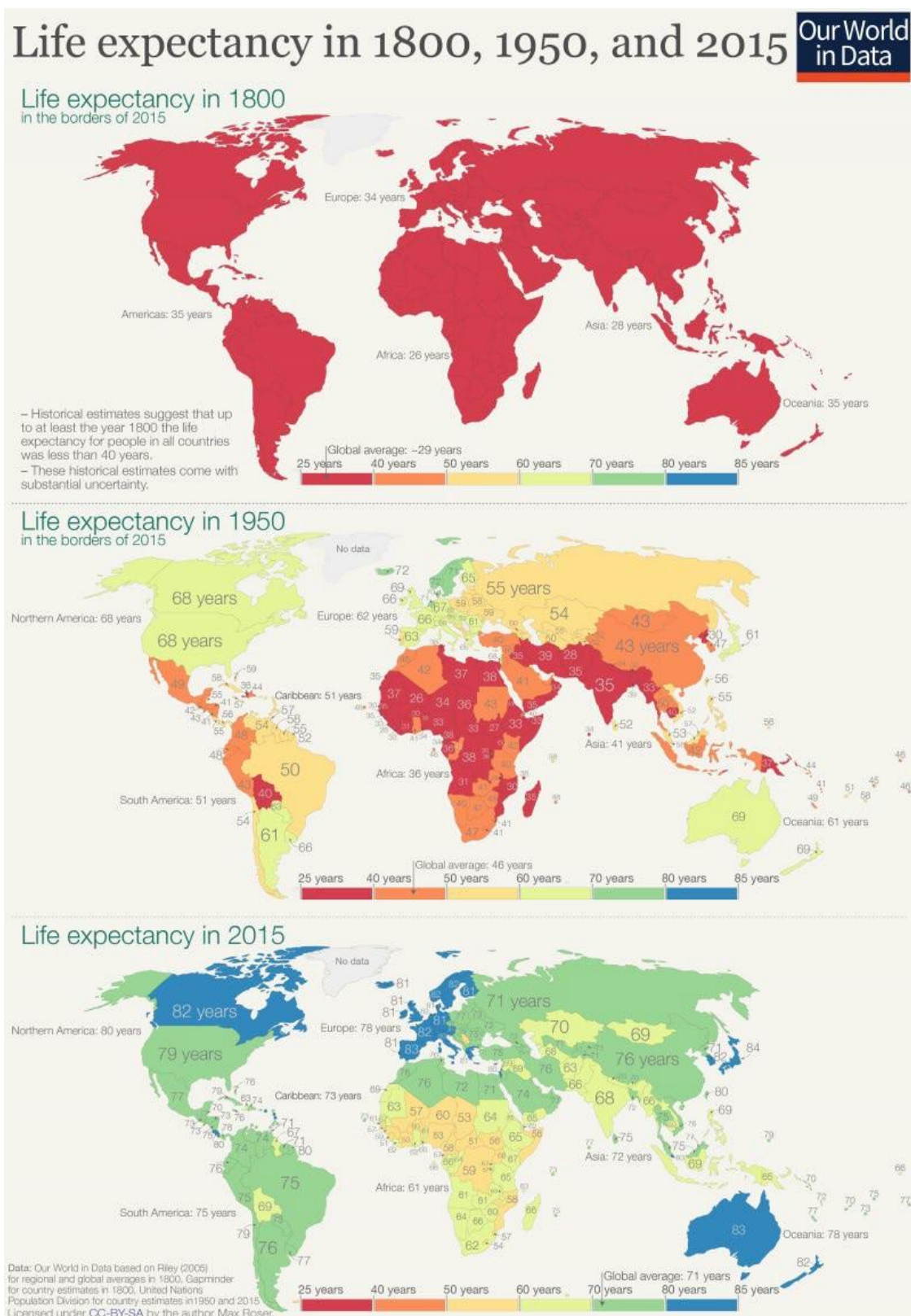
⁴⁰⁴ *Idem.*

⁴⁰⁵ Cfr. Jeff Desjardins, *The 50 Most Important Life-Saving Breakthroughs in History*, marzo 26, 2018, consúltese en: <https://www.visualcapitalist.com/50-important-life-saving-breakthroughs-history/> [15 de julio de 2022]

mientras que en Malí no pasaba de 26 años.⁴⁰⁶ La siguiente tabla nos muestra la esperanza de vida de los seres humanos en 1800, 1950 y 2015.

⁴⁰⁶ Max Roser, Esteban Ortiz-Ospina and Hannah Ritchie, *Life Expectancy*, octubre 2019, consúltese en: <https://ourworldindata.org/life-expectancy> [15 de julio de 2022]

Esperanza de vida de los seres humanos en 1800, 1950 y 2015



Fuente: *Life Expectancy*, octubre 2019 <https://ourworldindata.org/life-expectancy> [15 de julio de 2022]

Otro factor interesante que se debe tomar en cuenta para entender la evolución de los sistemas de salud es la relación que existe entre aquellas enfermedades que salieron del control del ser humano y los avances científicos y tecnológicos que poco a poco fueron apareciendo en la sociedad. En primer término, debemos decir que la humanidad, en múltiples ocasiones, ha sido afectada por pandemias y que este tipo de externalidades ha dejado lecciones importantes en para la mejora de la convivencia humana. Un factor que ha incrementado su margen de afectación está en el comercio y los medios de transporte (terrestres, marítimas y aéreas); ambos generaron las oportunidades para su dispersión a todas partes del mundo. Enfermedades como la tuberculosis, viruela, influenza fueron afectando a la población por la vía del transporte. Entre las principales pandemias que han ocurrido están las siguientes:

Principales pandemias en el mundo

Nombre	Periodo de tiempo	Tipo / huésped prehumano	Número de muertos
Plaga de Justiniano	541-542	Bacteria Yersinia pestis / Ratas, pulgas	30-50M
Muerte negra	1347-1351	Bacteria Yersinia pestis / Ratas, pulgas	200M
Gripe rusa	1889-1890	Se cree que es H2N2 (origen aviar)	1 M
Gripe española	1918-1919	Virus H1N1 / Cerdos	40-50M
Gripe asiática	1957-1958	Virus H2N2	1,1 M
Gripe de Hong Kong	1968-1970	Virus H3N2	1 M
VIH / SIDA	1981-presente	Virus / chimpancés	25-35M
La gripe porcina	2009-2010	Virus H1N1 / Cerdos	200,000
SARS	2002-2003	Coronavirus / Murciélagos, Civetas	770,000
Ébola	2014-2016	Ébola virus / Animales salvajes	11,000

Nombre	Periodo de tiempo	Tipo / huésped prehumano	Número de muertos
MERS	2015-actualidad	Coronavirus / Murciélagos, camellos	850,000
COVID-19	2019-actualidad	Coronavirus (Desconocido)	3, 128,962 (al 27 de abril de 2021)

Fuente: El autor, con base en: Nicolás Le Pan, *History of pandemics*, Visual Capitalist, <https://www.visualcapitalist.com/history-of-pandemics-deadliest/> y <https://covid19.who.int/> M = Millones. [15 de julio de 2022]

Si se hace una revisión histórica de los avances científicos y tecnológicos más importantes que han salvado millones de vidas y contribuido al modelo de salud actual, podemos encontrar que las vacunas, los antibióticos y las transfusiones de sangres han contribuido enormemente para detener los fallecimientos de las personas (algunos también incluyen a los inodoros). Lo que vemos hoy tan cotidiano, en otros momentos era imposible y, en conjunto, la higiene, la desinfección y la eliminación de bacterias mediante los antibióticos han cambiado al cuerpo del hombre y su entorno. Los avances de la medicina preventiva y de salud pública en los países en desarrollo son un parámetro para medir la transformación social del mundo.

Parte del cambio que se está dando en la medicina del siglo XXI está relacionado con el uso de las computadoras, el internet, los SED y la forma en la que trabaja el personal médico. Los medios tradicionales para almacenar los registros de los pacientes están siendo reemplazados por una memoria de computadora. Los médicos, ya sea en atención primaria o en hospitales, solicitarán la información para procesarla en computadoras. Las notas de casos y resúmenes están digitalizándose por norma; también las radiografías, las exploraciones corporales y otros datos visuales que se almacenen en sistemas expertos y está disponible instantáneamente en equipos del hospital en el que se podrá observar por el médico y, en su caso, el paciente. Los diagnósticos y tratamientos se están realizando con diversos equipos tecnológicos. Las pruebas de sangre (la glucosa y el colesterol), ya están disponibles en kits para uso directo de los pacientes. El desarrollo de biosensores electrónicos para monitorizar fluidos corporales en busca

de signos de mala salud se amplía en cobertura. Todo en conjunto estará influyendo en los modelos de salud que se tienen y aquellos a los que se aspira. La siguiente tabla nos muestra algunos avances científicos y tecnologías que han salvado vidas desde el siglo XVI a la fecha:

Avances científicos y tecnológicos que han salvado vidas

Año	Aportación tecnológica para el mejoramiento de la salud
1590	Se crea el primer microscopio
1816	Se inventa el estetoscopio
1850	Anestesia
1875	Toilettes
1890	Pasteurización
1895	Primer uso documentado de rayos X
1948	Desarrollo de lentes de contacto de plástico
1913	Transfusiones de sangre
1919	Cloración del agua
1922	Insulina
1928	Antibióticos
1945	Pesticidas
1955	Vacunas / Ultrasonido utilizado para embarazadas
1958	Se implanta el primer marcapasos cardíaco
1962	Primer reemplazo de cadera artificial moderno
1965	Se desarrolla la primera unidad de mamografía dedicada
1973	Se inventa la tomografía computarizada de todo el cuerpo
1970	Radiología
1977	Primera imagen de un ser humano en un escáner de resonancia magnética de cuerpo entero
1982	Primer corazón artificial permanente
1987	Primera cirugía con láser en una córnea humana
1995	Cirugía ocular Lasik aprobada por la FDA / Cirugía robótica
2000	Primer sistema robótico para cirugía laparoscópica general aprobado por la FDA
2003	Nanotecnología
2004	Escáner de tomografía computarizada de 64 cortes aprobado por la FDA
2016	Primer ensayo en humanos que utiliza la edición de genes CRISPR aprobado por los NIH
2021	Watson Health / Microsoft Health / Google Health / Amazon Health

Fuente: Elaboración propia con datos de: Imán Ghosh, Visualizing the Healthtech Revolution, en Visual Capitalist, marzo 26, 2019, consúltese: <https://www.visualcapitalist.com/visualizing-healthtech-revolution/> y Inventions that drive modern life expectancy, aperioncare, en <https://aperioncare.com/blog/inventions-life-expectancy/> The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database <https://www.nature.com/articles/s41746-020-00324-0> (DESARROLLAR)

El siglo XXI llega con la Medicina Basada en Evidencia (MBE) o “Evidence-based medicine”⁴⁰⁷, dejando muy atrás los postulados del modelo hipocrático tradicional⁴⁰⁸ contenidos en los manuales de medicina y cuyos cimientos traían ingredientes de las obras paradigmáticas como *Sobre la naturaleza del hombre*⁴⁰⁹ y los *Tratados*

⁴⁰⁷ El uso de este término fue acuñado en University de Hamilton, Canadá, 1992 por Gordon Guyatt (1992). Cfr. “An oral history of EBM” publicado por el Dr. Richard Smith en las revistas *Journal of the American Medical Association (JAMA)* y *British Medical Journal (BMJ)*, consúltese: [Evidence-Based Medicine: An Oral History \(jamanetwork.com\)](#) Guyatt G, Cairns J, Churchill D, et al. *Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine*. JAMA. 1992; 268(17):2420–2425. doi:10.1001/jama.1992.03490170092032

El *BMJ* publicó su primera edición el 3 de octubre de 1840 como *Provincial Medical and Surgical Journal* antes de unirse con *London Journal of Medicine* y publicarse a partir de enero de 1853 como *Associated Medical Journal*. Cuatro años más tarde, en enero de 1857, esta revista fusionada se convirtió en *British Medical Journal* y luego se cambió a *The BMJ* (2014). La *JAMA* se publica desde 1883 con investigaciones innovadoras en medicina.

⁴⁰⁸ El juramento hipocrático dice: “Juro por Apolo médico, por Asclepio, Higía y Panacea, así como por todos los dioses y diosas, poniéndolos por testigos, dar cumplimiento en la medida de mis fuerzas y de acuerdo con mi criterio a este juramento y compromiso: Tener al que me enseñó este arte en igual estima que a mis progenitores, compartir con él mi hacienda y tomar a mi cargo sus necesidades si le hiciere falta; considerar a sus hijos como hermanos míos y enseñarles este arte, si es que tuvieran necesidad de aprenderlo, de forma gratuita y sin contrato; hacerme cargo de la preceptiva, la instrucción oral y todas las demás enseñanzas 2 de mis hijos, de los de mi maestro y de los discípulos que hayan suscrito el compromiso y estén sometidos por juramento a la ley médica, pero a nadie más. Haré uso del régimen dietético para ayuda del enfermo, según mi capacidad y recto entender: del daño y la injusticia le preservaré. No daré a nadie, aunque me lo pida, ningún fármaco letal, ni haré semejante sugerencia. Igualmente, tampoco proporcionaré a mujer alguna un pesario abortivo. En pureza y santidad mantendré mi vida y mi arte. No haré uso del bisturí ni aun con los que sufren del mal de piedra: dejaré esa práctica a los que la realizan. A cualquier casa que entrare acudiré para asistencia del enfermo, fuera de todo agravio intencionado o corrupción, en especial de prácticas sexuales con las perso-nas, ya sean hombres o mujeres, esclavos o libres. Lo que, en el tratamiento, o incluso fuera de él, viere u oyere en relación con la vida de los hombres, aquello que jamás deba trascender, lo callaré teniéndolo por secreto. En consecuencia, séame dado, si a este juramento fuere fiel y no lo quebrantare, el gozar de mi vida y de mi arte, siempre celebrado entre todos los hombres. Mas si lo trasgredo y cometo perjurio, sea de esto lo contrario.”. Cfr. *Tratados hipocráticos*, Tomo I, Gredos, Madrid, 1983, pp. 65-66.

⁴⁰⁹ *Sobre la naturaleza del hombre* es un texto que vincula a la Filosofía con la medicina. La teoría humoral refleja la huella de Empédocles, “apunta directamente contra el monismo para abogar por una concepción pluralista de la *phýsis*.”. Este texto se encuentra dentro del *Corpus Hippocraticum*, que comprende más de 53 tratados y 72 libros sobre la profesión y Ética del médico, estudios de fisiología, patología, dietética y ginecología. “Los escritos más significativos... fueron compuestos entre 420 y 350 a. C., en la etapa que podemos considerar definitiva en la formación de la doctrina hipocrática... El médico que, según el elogio homérico (*Il.* XI 514), es “un hombre que vale por muchos otros”, busca actuar de manera consciente y metódica, confiando en la inquisición racional, conjugando su actividad técnica con una concepción amplia sobre los procesos naturales que afectan al ser humano como parte integrante de ese cosmos natural, regido por una *phýsis* universal.”. Ese concepto heredado de la Filosofía presocrática influye decisivamente en los escritores hipocráticos “... que unen sus dotes de observación minuciosa una capacidad notable de teorización sobre el hombre y el mundo”. Cfr. Hipócrates, *Tratados hipocráticos*, Tomos I (Introducción) y VIII (Introducción), Gredos, Madrid 1983.

hipocráticos (con soporte en la teoría humoral)⁴¹⁰, pero mantenemos el método heurístico para el diagnóstico y tratamiento de las enfermedades. Pareciera algo contradictorio, pero no lo es; queda confirmado que el cuidado del paciente debe mantenerse con este ingrediente y añadirle el uso de tecnología, el uso del método científico y las pruebas necesarias para el diagnóstico y tratamiento. La Medicina Basada en la Evidencia (MBE):

“...precisa la integración de los mejores indicios obtenidos a través de la investigación, nuestros conocimientos teóricos y prácticos y los valores y circunstancias específicos de cada paciente. • Con la expresión *mejores indicios obtenidos a través de la investigación* nos referimos a las investigaciones de interés clínico, que en ocasiones proceden de las ciencias básicas de la medicina, pero sobre todo de las investigaciones clínicas centradas en el paciente para determinar la exactitud y precisión de las pruebas diagnósticas (incluida la exploración clínica), la potencia de los marcadores pronósticos y la eficacia y seguridad de las estrategias terapéuticas, rehabilitadoras y preventivas. • Por *conocimientos teóricos y prácticos* entendemos la capacidad para utilizar nuestras habilidades clínicas y la experiencia adquirida para identificar rápidamente el estado de salud y el diagnóstico específicos de cada paciente, los riesgos y los beneficios derivados de las posibles intervenciones/ exposiciones/ pruebas diagnósticas en cada caso concreto y sus valores y expectativas personales. Además, son necesarios estos conocimientos clínicos para integrar la evidencia con los valores y las circunstancias del paciente. • Los *valores del paciente* son las preferencias, preocupaciones y expectativas únicas que cada paciente aporta a un encuentro clínico y que deben integrarse en las decisiones clínicas compartidas para que puedan ser de ayuda al paciente; por circunstancias del paciente nos referimos a la situación clínica individual del paciente y al entorno clínico.”⁴¹¹

La Universidad de Oxford, considerada el espacio académico número uno en ciencias médicas, creó el Centro de Medicina Basada en Evidencia (CEBM) en el que se investiga, practica y difunde este tipo de conocimiento. El CEBM está

⁴¹⁰ La naturaleza humana es explicable a través de sus humores y estos constituyen el *Phýon* humano -sangre, pituita, bilis amarilla y bilis negra- (agua, fuego, aire y tierra, previamente). Su equilibrio es el origen de la salud o enfermedad. La desaparición de alguno o la alteración de alguno produce la enfermedad, “ya por la evacuación de una determinada cantidad fuera del cuerpo (*kénosis*), ya por el desplazamiento del humor en su interior (*metástasis*) o bien por el establecimiento del humor en un lugar del cuerpo que no le corresponde por naturaleza (*apóstasis*). Los cuatro humores... están sujetos a la influencia de las estaciones, en un tradicional intento de establecer vínculos entre el macrocosmos y el microcosmos... las cuatro estaciones del año se corresponden con las cuatro cualidades generales: lo frío, lo cálido, lo húmedo y lo seco. Así se establece ciclo continuo pituita, fría y seca, en invierno; la sangre, cálida y húmeda, en primavera; la bilis amarilla, cálida y seca, en verano; finalmente, la bilis negra, fría y húmeda, en otoño.” Cfr. *Tratados hipocráticos*, Tomo VIII, Gredos, Madrid 1983, Introducción.

⁴¹¹ Straus, Sharon E. *Medicina basada en la evidencia*, Elsevier Health Sciences. Edición de Kindle, Madrid, 2019, pp. 1-2.

aprovechando los avances de la tecnología (brazos robóticos, órganos bioimpresos en 3D, realidad virtual para aliviar el dolor, Internet de las cosas (IoT) y la salud móvil (mHealth) para ofrecer a médicos y pacientes una forma conveniente de reservar citas y pagar facturas en línea, hasta permitir que los médicos utilicen registros médicos electrónicos para acceder y compartir información. Los dispositivos portátiles y las aplicaciones de teléfonos inteligentes están aumentando su adopción a medida que desbloquean la opción de monitorear y administrar la salud individual en cualquier momento y en cualquier lugar. Esto está creando una explosión en los datos de salud personal, que los consumidores están dispuestos a compartir con su médico si los beneficiará a ellos o a otros. Si bien la tecnología de la salud no reemplazará al médico, sí lleva a un cambio en la experiencia del paciente.

Las evidencias en el ámbito de la medicina ayudan a dar soporte a las afirmaciones y convierte a las investigaciones en herramientas confiables para la toma de decisiones de calidad. Durante los últimos 20 años Oxford ha construido un equipo de investigación multidisciplinario comprometido con la promoción de los principios de la MBE.

Ma. Eugenia Ponce de León C., secretaria de Educación Médica, de la Facultad de Medicina de la UNAM, comenta que la práctica de la MBE consta de una secuencia de 5 pasos que se deben seguir para ordenar el pensamiento y el actuar de los profesionales de la salud:

1. Transformar la necesidad de información (sobre diagnóstico, tratamiento, pronóstico o daño) en una pregunta clara y precisa.
2. Buscar la mejor evidencia, que dé respuesta a la pregunta formulada.
3. Valorar críticamente la evidencia encontrada, en cuanto a su validez, impacto y utilidad.
4. Incorporar esa evaluación crítica a nuestra experiencia clínica, al problema específico y a las circunstancias del paciente.

5. Evaluar nuestra efectividad y eficiencia en la ejecución de los cuatro pasos anteriores y buscar mejorar la ejecución de los mismos para las siguientes búsquedas.⁴¹²

⁴¹² Ponce de León C., Ma. Eugenia, Enseñanza de la medicina basada en la evidencia, Revista de la Facultad de Medicina, UNAM, Vol. 44, número 3, mayo-junio 2001, p. 125.

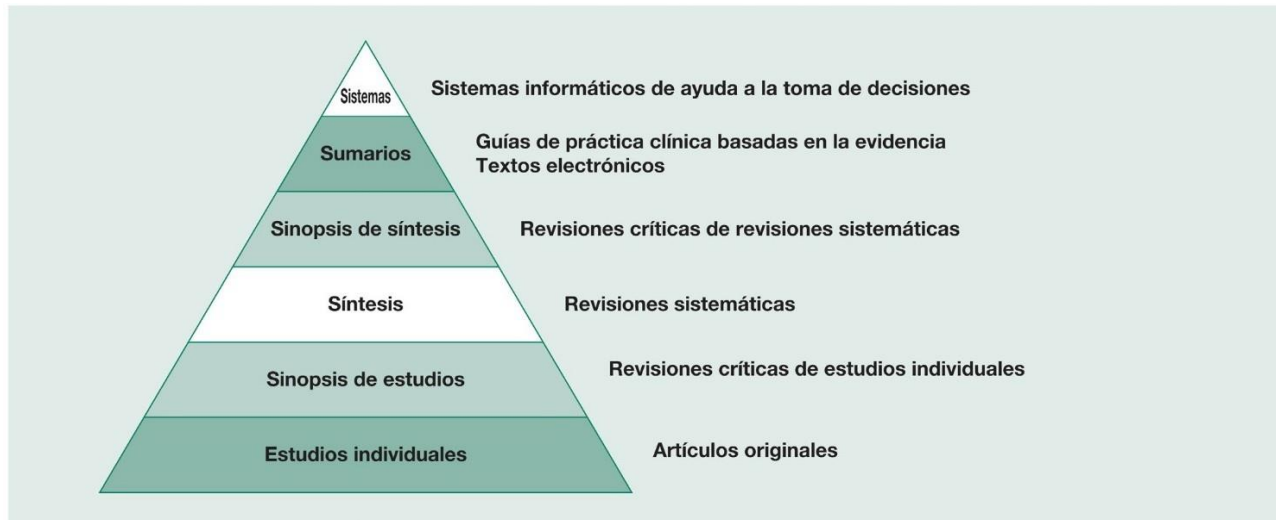
Formulación de una pregunta clínica		
Elemento de la pregunta	Guía para la formulación	Ejemplo
P Paciente/problema de salud	¿Cómo podría describir a un grupo de pacientes similares al mío?	<i>¿En los sujetos diabéticos tipo 2...</i>
I Intervención de estudio	¿Qué intervención principal (tratamiento, posible causa, prueba diagnóstica, factor pronóstico) estoy considerando?	<i>... la administración de antiagregantes plaquetarios...</i>
C Intervención de comparación	¿Cuál es la principal alternativa para comparar la intervención?	<i>... en relación con no administrarlos...</i>
O Variable de resultado (outcome)	¿Qué deseo conseguir? ¿Qué resultados de la intervención pretendo valorar? ¿Cuál es la variable que puede evaluar el resultado de interés clínico?	<i>... disminuye el riesgo de mortalidad y morbilidad cardiovascular?</i>

Fuente: J. Ma. Argimón Pallàs, A. Guarga Rojas, J. Jiménez Villa, Medicina basada en la evidencia: guías y protocolos, descargado por la Universidad Nacional Autónoma de México, de ClinicalKey.es por Elsevier, junio 25, 2020, pp. 72-91, consúltase en: <http://dsp.facmed.unam.mx/wp-content/uploads/2016/06/U-2-Anexo-1A-Argim%C3%B3n-J-Medicina-Basada-en-evidencias-gu%C3%ADas-y-protocolos.pdf> [15 de julio de 2022]

El segundo paso, es decir, la búsqueda de evidencia apropiada para responder a la pregunta ocupa del uso de fuentes de información, obligadamente, de gran variedad y cantidad, que a su vez requiere de una clasificación jerárquica, denominada por algunos académicos estructura 6S. El profesional de la salud iniciará la búsqueda en el recurso de mayor nivel disponible para el problema, con la finalidad de aprovechar el trabajo ya realizado por otros investigadores y con ello ahorrar tiempo en la búsqueda de datos y en la obtención de resultados.⁴¹³ Existen bases de datos para búsquedas específicas como *National Guidelines Clearinghouse* (www.guideline.gov) o la *Guidelines International Network* (www.g-i-n.net); *GuiaSalud*: (www.guiasalud.es); *MeSH*, *Medical Subject Headings* (www.ncbi.nlm.nih.gov/mesh). La siguiente tabla nos muestra los seis niveles de la estructura 6S:

413

Niveles de organización de la evidencia: estructura 6S.



Fuente: J. Ma. Argimón Pallàs, A. Guarga Rojas, J. Jiménez Villa, Medicina basada en la evidencia: guías y protocolos, descargado por la Universidad Nacional Autónoma de México, de ClinicalKey.es por Elsevier, junio 25, 2020, pp. 72-91, consúltese en: <http://dsp.facmed.unam.mx/wp-content/uploads/2016/06/U-2-Anexo-1A-Argim%C3%B3n-J-Medicina-Basada-en-m-evidencias-gu%C3%A1s-y-protocolos.pdf> [15 de julio de 2022]

En la práctica cotidiana los profesionales de la salud toman múltiples decisiones sobre la atención del paciente, relacionadas con la prevención, diagnóstico, pronóstico, tratamiento o la gestión de recursos sanitarios en las que la incertidumbre está presente a cada momento. El diagnóstico de un médico puede cambiar con los mismos elementos cuando otro médico interviene o cuando el mismo médico conoce y con el tiempo modifica su postura. Este proceso de toma de decisiones es natural en la práctica clínica. J. Ma. Argimón Pallàs, A. Guarga Rojas, J. Jimenez Villa, señalan que:

Las variaciones de la práctica clínica pueden estar asociadas a factores intrínsecos del proceso asistencial, como la complejidad de la anamnesis, las dificultades para predecir el curso de una enfermedad o para interpretar los hallazgos exploratorios, o la multiplicidad de opciones diagnósticas y terapéuticas. Pero la dificultad para identificar la mejor alternativa y tomar la decisión clínica más adecuada depende también de factores extrínsecos, como los resultados clínicos previsibles, el coste de la asistencia, los recursos disponibles, el impacto familiar y social, o los valores culturales. A todo ello se añade el aumento de la complejidad clínica asociado a la pluripatología, la imposibilidad de acceder y analizar en tiempo real toda la información científica, la necesidad de gestionar los recursos sanitarios de forma eficiente y equitativa, las creencias y los valores de los profesionales o la aparición de un nuevo

modelo de usuario que exige participar de forma activa en el proceso de toma de decisiones de salud. El importante desarrollo tecnológico actual, tanto en el ámbito de la información y la comunicación como en el propiamente sanitario, junto con la globalización del acceso a la información, ha producido un importante incremento de las expectativas, tanto de los pacientes y la comunidad como de los propios profesionales, y ha conducido a la exigencia de una atención de la máxima calidad posible. De hecho, existe una presión creciente sobre los profesionales y el sistema sanitario para que las decisiones se basen en la mejor evidencia científica disponible y no solamente en las opiniones de determinados expertos o grupos de opinión. El profesional sanitario puede tomar las decisiones por tradición (“Siempre se ha hecho así”), por convención (“Todos lo hacen así”), por creencias o dogma (“Creo que hacerlo así es mejor”), o bien basándose en pruebas científicas. En este último caso, debe integrar tres elementos: las mejores pruebas científicas disponibles, las características, necesidades, valores y preferencias del paciente, y los recursos disponibles, incluyendo las habilidades y la experiencia del profesional. Además, es totalmente imprescindible enmarcar estas decisiones dentro del contexto y de la organización en que se toman. Cada uno de estos elementos desempeña un papel crucial, y por ello es imprescindible disponer de información fiable y válida sobre cada uno de ellos y considerar cómo aplicarla a la situación concreta. El reconocimiento de la necesidad de que las decisiones se basen en las mejores pruebas disponibles procedentes de la investigación científica, identificadas y evaluadas con rigor, ha hecho que, en los últimos años, los profesionales sanitarios muestren un interés creciente en los aspectos relacionados con la denominada “medicina basada en la evidencia” (MBE).⁴¹⁴

El CEBM de Oxford tiene varios proyectos de MBE, entre los que destacan:

- 1) COVID-19 (Revisiones rápidas de evidencia, análisis de datos y escritos que invitan a la reflexión sobre la pandemia de coronavirus, actualizados regularmente por nuestro equipo con colaboradores internacionales);
- 2) Transmisión de SARS-CoV-2 (proporciona un resumen y una evaluación rápidos de los datos relevantes en torno a la transmisión del SARS-CoV-2, informa sobre las implicaciones políticas importantes y destaca las áreas de investigación que se necesitan con urgencia).;
- 3) Cochrane (cigarrillos electrónicos para dejar de fumar);
- 4) Uso de anticoagulantes;
- 5) Primodos y malformaciones congénitas;
- 6) Daños de la malla

⁴¹⁴ J. Ma. Argimón Pallàs, A. Guarga Rojas, J. Jiménez Villa, Medicina basada en la evidencia: guías y protocolos, descargado por la Universidad Nacional Autónoma de México, de ClinicalKey.es por Elsevier, junio 25, 2020, pp. 72-91, consúltase en: <http://dsp.facmed.unam.mx/wp-content/uploads/2016/06/U-2-Anexo-1A-Argim%C3%B3n-J-Medicina-Basada-en-m-evidencias-gu%C3%ADas-y-protocolos.pdf> [15 de julio de 2022]

transvaginal.⁴¹⁵ A continuación, explicaremos la llamada *Healthcare technology* y su importancia para el mundo actual.

⁴¹⁵ Para entender más sobre la MBE consúltase: Nunan D, O'Sullivan J, Heneghan C, *et. Al.*, Ten essential papers for the practice of evidence-based medicine, *BMJ Evidence-Based Medicine*, febrero 2018; en <https://ebm.bmj.com/content/22/6/202> [15 de julio de 2022]

A) Healthcare technology

La *Healthcare technology* o *Health technology* es definida por la Organización Mundial de la Salud (OMS) como “la aplicación de conocimientos y habilidades organizados en forma de dispositivos, procedimientos, sistemas, medicamentos y vacunas desarrollados para salvar un problema de salud y mejorar la calidad de vida”.⁴¹⁶

La Organización Mundial de la Salud, reconoce que las tecnologías sanitarias dotan a los proveedores de servicios de salud de herramientas que son necesarias para la prevención, el diagnóstico, el tratamiento y la rehabilitación (efectivos y eficientes) y el logro de las metas que en materia de salud se tienen programadas para los próximos años.

⁴¹⁶ En 2007, la LX Asamblea Mundial de la Salud, se reconoció que las tecnologías sanitarias ofrecen a los prestadores de servicios de salud herramientas necesarias para la prevención, el diagnóstico, el tratamiento y la rehabilitación efectivos y eficientes. Entendiendo que las tecnologías sanitarias, en particular los dispositivos médicos, representan un desafío económico y técnico para los sistemas de salud de muchos estados, y preocupado por el desperdicio de recursos resultante de inversiones inapropiadas en tecnologías sanitarias, en particular dispositivos médicos, que no satisfacen necesidades de alta prioridad, son incompatibles con las infraestructuras existentes, se utilizan de forma irracional o incorrecta, o no funcionan de manera eficiente; Existe una necesidad de que los estados y los donantes frenen los crecientes costos estableciendo prioridades en la selección y adquisición de tecnologías sanitarias, en particular de dispositivos médicos, sobre la base de su impacto en la carga de morbilidad, y de asegurar el uso eficaz de los recursos a través de planificación, evaluación, adquisición y gestión; Se requiere ampliar los conocimientos especializados en el campo de las tecnologías sanitarias, en particular los dispositivos médicos; La OMS pide a sus miembros: 1) recopilar, verificar, actualizar e intercambiar información sobre tecnologías sanitarias, en particular dispositivos médicos, como ayuda para la priorización de necesidades y la asignación de recursos; 2) formular, según proceda, estrategias y planes nacionales para el establecimiento de sistemas de evaluación, planificación, adquisición y gestión de tecnologías sanitarias, en particular dispositivos médicos, en colaboración con el personal involucrado en la evaluación de tecnologías sanitarias y la Ingeniería biomédica; 3) elaborar directrices nacionales o regionales de buenas prácticas de fabricación y reglamentarias, establecer sistemas de vigilancia y otras medidas para garantizar la calidad, seguridad y eficacia de los productos sanitarios y, en su caso, participar en la armonización internacional; (4) establecer, cuando sea necesario, instituciones regionales y nacionales de tecnología sanitaria, y colaborar y crear asociaciones con los proveedores de atención de la salud, la industria, las asociaciones de pacientes y las organizaciones profesionales, científicas y técnicas; (5) recopilar información que interrelacione los dispositivos médicos, que se ocupen de las condiciones prioritarias de salud pública en diferentes niveles de atención y en diversos entornos y entornos, con la infraestructura, los procedimientos y las herramientas de referencia necesarios. Cfr. World Health Organization, WHA60.29 Health technologies, Sixtieth World Health Assembly, Ginebra, 14 a 23 mayo de 2007, Resolutions and Decisions Annexes, consul
tese en: https://www.who.int/healthsystems/WHA60_29.pdf?ua=1 [15 de julio de 2022]

En particular los dispositivos médicos, representan un desafío económico y técnico para los sistemas de salud de muchos Estados miembros por la brecha tecnológica que existe, los riesgos de corrupción y su mal uso ya que en muchas contrataciones se ha presentado un desperdicio de recursos por inversiones inapropiadas en tecnologías sanitarias, que no satisfacen necesidades ni prioridades internas de los países o que son incompatibles con las infraestructuras existentes, utilizándose de forma incorrecta e ineficiente.

Una solución que plantea la OMS tiene que ver con la reducción de los costos de la adquisición de dispositivos médicos y la ubicación de las prioridades en la selección y adquisición de tecnologías sanitarias, sobre la base de su impacto en la carga de morbilidad, y de asegurar el uso eficaz de los recursos a través de planificación, evaluación, adquisición y gestión adecuada. La OMS recomienda:

(1) to collect, verify, update and exchange information on health technologies, in particular medical devices, as an aid to their prioritization of needs and allocation of resources; (2) to formulate as appropriate national strategies and plans for the establishment of systems for the assessment, planning, procurement and management of health technologies, in particular medical devices, in collaboration with personnel involved in health-technology assessment and biomedical engineering; (3) to draw up national or regional guidelines for good manufacturing and regulatory practices, to establish surveillance systems and other measures to ensure the quality, safety and efficacy of medical devices, and where appropriate to participate in international harmonization; (4) to establish where necessary regional and national institutions of health technology, and to collaborate and build partnerships with health-care providers, industry, patients' associations and professional, scientific and technical organizations; (5) to collect information that interrelates medical devices, which deal with priority public-health conditions at different levels of care and in various settings and environments, with the required infrastructure, procedures and reference tools; [...]⁴¹⁷

⁴¹⁷ 1) recopilar, verificar, actualizar e intercambiar información sobre tecnologías sanitarias, en particular dispositivos médicos, como ayuda para la priorización de necesidades y la asignación de recursos; (2) formular, según proceda, estrategias y planes nacionales para el establecimiento de sistemas de evaluación, planificación, adquisición y gestión de tecnologías sanitarias, en particular dispositivos médicos, en colaboración con el personal involucrado en la evaluación de tecnologías sanitarias y la Ingeniería biomédica; (3) elaborar directrices nacionales o regionales de buenas prácticas de fabricación y reglamentarias, establecer sistemas de vigilancia y otras medidas para garantizar la calidad, seguridad y eficacia de los productos sanitarios y, en su caso, participar en la armonización internacional; (4) establecer, cuando sea necesario, instituciones regionales y nacionales de tecnología sanitaria, y colaborar y crear asociaciones con los proveedores de atención de la salud, la industria, las asociaciones de pacientes y las organizaciones profesionales, científicas y técnicas; (5) recopilar información que interrelacione los dispositivos médicos, que se ocupen de

Las empresas de tecnología más poderosas, como Amazon, IBM, Alphabet, Microsoft, Apple y Google están intentando brindar su soporte tecnológico a los servicios de salud en gran parte del mundo. Datos que están generándose en la red nos muestran la relevancia de que las empresas de tecnología participen en este sector. Por ejemplo:

- En los EE. UU., Habrá 98,2 millones de personas mayores de 65 años para 2060
- La diabetes afectará a 642 millones de personas en todo el mundo para 2040
- El 70% de las empresas de atención médica están invirtiendo en tecnología orientada al consumidor, como aplicaciones, monitoreo remoto y atención virtual.
- La tecnología de monitoreo remoto de pacientes podría ahorrarle al sistema de salud \$ 200 mil millones durante los próximos 25 años
- Para el 2030, 171 millones de estadounidenses tendrán alguna forma de enfermedad crónica y Estados Unidos enfrentará una escasez de hasta 104,900 médicos.
- El auge de los teléfonos inteligentes y su crecimiento en todo el mundo está cambiando la experiencia del mercado de la salud.
- El 59% de los consumidores esperan que sus servicios de atención médica sean tan buenos como los de Amazon.
- El modelo de atención médica tradicional ya no se alinea con la mentalidad del consumidor. Se han identificado seis fuerzas transformadoras de la atención médica: 1) Los datos; 2) Flujo de información en registros de salud electrónicos; 3) Los consumidores piden la información de salud en un formato

las condiciones prioritarias de salud pública en diferentes niveles de atención y en diversos entornos y entornos, con la infraestructura, los procedimientos y las herramientas de referencia necesarios; [...] Cfr. World Health Organization, WHA60.29 Health technologies, Sixtieth World Health Assembly, Ginebra, 14 a 23 mayo de 2007, Resolutions and Decisions Annexes, consúltese en: https://www.who.int/healthsystems/WHA60_29.pdf [15 de julio de 2022]

- personalizado, accesible, atractivo y entendible; 4) Los consumidores quieren que la industria de la salud los encuentre, los conozca y los ayude, utilizando comportamientos pasados para anticipar sus necesidades futuras; 5) Las nuevas tecnologías están sirviendo para reclutar pacientes adecuados para ensayos clínicos, reduciendo ineficiencias en el reclutamiento; 6) Hay una transformación del modelo de ventas
- La Organización Mundial de la Salud estima que hay una escasez de más de 5 millones de médicos en todo el mundo, lo que deja a más de la mitad de la población mundial sin acceso ni siquiera a los servicios de salud más básicos.⁴¹⁸

Derivado de lo anterior, las grandes empresas están guiando las tendencias digitales en salud con varios productos y servicios relacionados con: realidad virtual, telemedicina, wearables, hospitales inteligentes, nanotecnología, inteligencia artificial, entre otros. La siguiente tabla nos muestra estas tendencias digitales en salud:

⁴¹⁸ Marc Doucette, Why Big Tech is Plotting an Invasion of the Healthcare Market, Visual capitalist, abril 24, 2018, consúltese en: <https://www.visualcapitalist.com/big-tech-invading-healthcare-market/>; The 6 Forces Transforming the Future of Healthcare, Visual capitalist, julio 12, 2018, consúltese en: <https://www.visualcapitalist.com/6-forces-transforming-future-healthcare/> y <https://www.onlinedoctor.com/healthcare-jobs-medical-technology/> <https://www.forbes.com/sites/adigaskell/2018/08/09/integrating-innovation-into-healthcare/?sh=156b0731632e> [15 de julio de 2022]

Tendencias digitales y salud

Tendencia	Beneficios/características
Realidad aumentada (virtual)	Herramienta de formación 3D, que ofrece a los estudiantes de medicina y a los médicos una experiencia de aprendizaje tangible sin necesidad de utilizar cadáveres. Agrega mejoras de visualización 3D a los escaneos médicos.
Telemedicina	Hace que los chequeos sean más accesibles y convenientes. Ahorra tiempo. Especialmente útil para la salud mental, donde a menudo se necesita ayuda de manera oportuna. A menudo más barato que la atención médica tradicional.
Teléfonos inteligentes	El 80% de los médicos ya utilizan teléfonos inteligentes y aplicaciones médicas, y el 72% accede a información sobre medicamentos en los teléfonos inteligentes de forma regular.
Wearables y aplicaciones sanitarias	Los propietarios de teléfonos inteligentes pueden controlar sus niveles de oxígeno, frecuencia cardíaca, presión arterial y temperatura, todo sin salir de casa. Libera la disponibilidad del médico. Más conveniente para los pacientes. Las mediciones se pueden informar al médico. Esto podría significar controles menos frecuentes y prevenir problemas antes de que surjan.
Hospitales inteligentes	Reducir errores. Predecir y prevenir problemas de salud. Brindar a los pacientes una atención más personalizada.
Impresión 3D	Los nuevos impresores pueden crear herramientas médicas utilizando cualquier material edificable, desde plástico hasta células madre: Huesos artificiales para reconstrucción quirúrgica; Prótesis a medida, aparatos ortopédicos para escoliosis ventilados, soportes para amputados; Órganos diminutos hechos de células madre en capas que pueden crecer dentro del cuerpo de un paciente y tomar el control cuando falla un órgano viejo
Cirugía por Robot	Los sistemas asistidos por robot más utilizados son las cámaras y los brazos mecánicos con herramientas quirúrgicas adjuntas. Estos brazos son controlados por un cirujano a través de una computadora cercana, donde ve una vista ampliada de HD del sitio quirúrgico.
Nanomedicina	La nanomedicina es el uso de nanotecnología y nano dispositivos para mejorar el tratamiento y el diagnóstico médico a una escala extremadamente pequeña. Los medicamentos más pequeños y los sistemas de administración más precisos permitirán a los médicos: Administrar quimioterapia directamente a los tumores específicos en lugar de envenenar todo el cuerpo. Reducir la frecuencia de las inyecciones de medicamentos liberando lentamente el medicamento de una sola nanopartícula. ⁴¹⁹
Tecnología en la nube	La nube permite gestionar datos para dar mejores diagnósticos y tratamientos y dar seguimiento al historial del paciente, así como para mantenerse al día con las investigaciones. ⁴²⁰ (IBM WATSON). Permite que médicos y pacientes se comuniquen y ayuda a pacientes a estar más sanos.

⁴¹⁹ Un nanómetro es la mil millonésima parte de un metro y hay 25,400,000 nanómetros en una pulgada.

⁴²⁰ <https://www.mbahealthcaremanagement.org/topics/how-technology-is-changing-healthcare/> [15 de julio de 2022]

Inteligencia artificial con ML	Los verificadores de síntomas, diagnósticos y tratamientos, en línea, tienen un potencial significativo para mejorar la atención al paciente gracias al <i>Machine Learning</i> . ⁴²¹

Fuente: El autor con base en <https://www.visualcapitalist.com/big-tech-invading-healthcare-market/> y <https://www.onlinedoctor.com/healthcare-jobs-medical-technology/> <https://www.mbahealthcaremanagement.org/topics/how-technology-is-changing-healthcare/>

Estas tendencias influirán y cambiarán la atención médica tradicional, la cual se irá desplazando hacia el uso de tecnología ya sea por regulación o por decisión propia. La MBE y la *Healthcare technology*, puede dar a los médicos herramientas más exactas para atender la problemática de los pacientes y utilizar los avances tecnológicos en sus prácticas cotidianas (no siempre rigurosas ni con soporte científico). La transformación de los datos clínicos en conocimiento científico válido, estadísticamente preciso y con resultados relevantes para atender al paciente. A lo largo de la historia de la humanidad, el oficio del médico “ha sido eminentemente empírico, basado en conocimientos anatómicos y fisiopatológicos, así como en el sentido común, pero con postulados no siempre contrastables o reproducibles... Aún hoy se calcula que el 20-30% de las prácticas habituales son empíricas y no se ha evaluado su fundamento científico. Las fuentes de autoridad tradicionales, basadas en el prestigio personal, el sentido común, la experiencia personal y profesional o la tradición, se han visto progresivamente desplazadas por la evidencia proporcionada por la aplicación rigurosa del método científico mediante la investigación”⁴²². A su vez, la tecnología está reduciendo las limitantes de tiempo y recursos (humanos, materiales y tecnológicos).

No se trata de desaparecer la intuición y el buen juicio clínico, tampoco el acotar las habilidades personales o las bases de la formación académica de los médicos ni mucho menos reducir la comunicación con el paciente. Se busca aportar

⁴²¹ Puede consultarse un estudio introductorio a este tema: *A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis*, en: <https://arxiv.org/pdf/1806.10698.pdf> [15 de julio de 2022]

⁴²² M. Vega-de Céniga , N. Allegue-Allegue, S. Bellmunt-Montoya, C. López-Espada, R. Riera-Vázquez, T. Solanich-Valldaura, J. Pardo-Pardo, *Medicina basada en la evidencia: concepto y aplicación*, Angiología, 2009, número 61, pp. 29-34, en: <https://www.elsevier.es/es-revista-angiologia-294-articulo-medicina-basada-evidencia-concepto-aplicacion-S0003317009110040> [15 de julio de 2022]

un ingrediente adicional a ese conjunto de factores que ya se han utilizado por varios siglos: el método científico y su metodología. Aprender de estas bases metodológicas que nos permitan encontrar lo necesario para un tratamiento que beneficie al paciente. Aprovechar el internet, el *big data*, el cómputo en la nube, el internet de las cosas, el *Machine Learning* o, en su caso, mantener el uso de los tradicionales sistemas expertos decisionales como una herramienta de asertividad médica.

La *Healthcare technology* está teniendo un gran desarrollo gracias a los nuevos smartphones, sus capacidades y la velocidad de las redes 4G y 5G (sobre todo esta última conjuntamente con el internet de las cosas). Pensemos tan solo todo lo que actualmente puede contener un dispositivo telefónico: cámara, grabadora de video, GPS, calculadora, reloj, despertador, reproductor de música, grabadora de voz, álbum de fotos. y biblioteca de libros. Y también una infinidad de aplicaciones bajo demanda de su usuario: desde el tradicional diccionario o traductor hasta aquellas que te permiten hacer ejercicio, cocinar, estudiar, reparar algo en tu casa o que seas atendido por un *bot* en algún centro hospitalario.

Cabe señalar que el sector médico, como el de los abogados, son altamente conservadores de tradiciones y de formalismos que se vienen utilizando desde siglos atrás, ambos se cosifican en sus reglas. Más allá de su modificación o de luchar contra la resistencia de ciertos sectores, las empresas que comercializan medicamentos y este tipo de dispositivos o pruebas para diagnósticos y las agencias reguladoras, derivado de la pandemia del COVID-19, están obligadas a tomar el control de estos temas y a normar su uso. Muchos de los cambios se irán a las legislaciones de cada país, otros ya están siendo estudiados por organismos supranacionales.

Imaginemos, por ejemplo, qué tendría que hacer un legislador ante actividades que ya se están dando como: 1) monitorear de manera remota y continua los latidos del corazón; 2) medir la presión arterial a cada momento, la frecuencia y la profundidad de la respiración, la temperatura corporal, la concentración de oxígeno en la sangre, la glucosa, las ondas cerebrales, la

actividad, el estado de ánimo; obtener imágenes de cualquier parte del cuerpo y hacer una reconstrucción tridimensional, lo que eventualmente conduce a la capacidad de imprimir un órgano; y, 3) usar un dispositivo de imágenes de alta resolución, portátil y en miniatura que captura rápidamente información crítica en cualquier lugar.⁴²³ Simplemente, no tendría ni las habilidades, ni al staff técnico para atender la regulación de este tipo de temas.

Organizaciones como la IEEE realizó la discusión de este tema para 2021 en su *Conference on Connected Health: Applications, Systems and Engineering Technologies* (CHASE) 16 al 17 de diciembre de 2021 en la Ciudad de Washington D.C., Estados Unidos.⁴²⁴ La IEEE, preocupada por este tema, busca reunir a investigadores que trabajan en el área de la salud y la tecnología para intercambiar ideas en temas torales para la ciencia médica como:

Conference on Connected Health: Applications, Systems and Engineering Technologies

Aplicaciones	Sistemas	Tecnologías de Ingeniería
- Salud pública y monitoreo del hogar - Procesamiento de datos de atención médica, análisis de datos, minería de datos - Imágenes médicas y Computación portátil - Medicina personalizada y tratamiento preventivo - Modelos evolutivos y longitudinales de pacientes y enfermedades - Seguridad, privacidad y confianza para aplicaciones / servicios de salud conectados	- Sistemas de monitoreo de sensores biomédicos - Integración de dispositivos médicos con salud inalámbrica - Redes de área corporal - Redes de sensores implantables - Sistemas de diagnóstico médico remoto - Diseño de sistemas de comunicación o de apoyo a la toma de decisiones	- Diseño de dispositivos portátiles - Prendas de vestir / textiles inteligentes - Comunicaciones / Infraestructuras y protocolos de red - Robótica para la salud conectada - Sistema de diagnóstico médico remoto - Software, sistemas e Ingeniería de rendimiento para la salud conectada

Fuente: El autor.

⁴²³ Caiotlin Cheadle, "Here's How Tech is Disrupting the Traditional Healthcare Market", *Visualcapitalist*, diciembre 13, 2016, en <https://www.visualcapitalist.com/tech-disrupting-healthcare-market/> y How Cloud Technology Is Changing Healthcare en: <https://www.mbahealthcaremanagement.org/topics/how-technology-is-changing-healthcare/> [15 de julio de 2022]

⁴²⁴ Puede consultar más datos del evento en: <https://conferences.computer.org/chase2021/index.html> [15 de julio de 2022]

No todo es sencillo como parece en la industria de la *Healthcare technology*. Existen varios obstáculos a los que se enfrentan sus promotores. En particular: la visión de las élites políticas, la influencia de diversos actores económicos que controlan el mercado, el incremento del envejecimiento poblacional, la prevalencia de enfermedades crónico-degenerativas y, sobre todo, la corrupción que acecha a todo.

En cuanto a la dinámica del comportamiento que están teniendo las empresas vinculadas a la tecnología y a la salud, se encontraron datos que señalan al año 2016 como un momento clave para las fusiones y adquisiciones de tecnología de salud digital. Se dieron 45 fusiones y adquisiciones de salud digital⁴²⁵. La siguiente tabla contempla algunas de las más importantes:

⁴²⁵ [Forty-five 2016 digital health mergers and acquisitions | MobiHealthNews](#) [15 de julio de 2022]

Fusiones y adquisiciones de empresas de tecnología para la salud digital

Fusión	Servicios/Productos
Imprivata aceptó ser adquirido por Thoma Bravo	Desarrollador de herramientas de seguridad de TI en salud
Evolent Health adquiere a Valence health.	Proporciona una plataforma de atención a los principales sistemas de salud y grupos médicos en los Estados Unidos, a través de servicios de soporte clínico, analítico y financiero
Nokia Technologies adquiere Withings	Nokia, ofrece soluciones de red privada para la atención médica en tiempo real y la nube Withings, construye básculas para teléfonos inteligentes, puños de presión arterial, rastreadores de actividad, termómetros
DocPlanner se fusiona con Doctoralia	Ofrece a los profesionales de la salud y centros médicos herramientas que les ayudan a gestionar a sus pacientes, mejorar la eficiencia de su consulta y potenciar su presencia online
Royal Philips adquirió Wellcentive	una compañía de software para gestión de servicios de salud
Zipongo adquirió FillMyFork	Una aplicación de recompensas orientada a los empleados que promueve una alimentación saludable
Apple adquirió el startup Glimpse	Generación de datos médicos

Fuente: El autor con base en: [Forty-five 2016 digital health mergers and acquisitions | MobiHealthNews](#) [15 de julio de 2022]

Existen seis factores que los expertos han detectado y que se deben tomar en cuenta por las empresas *Healthcare technology* para configurar su participación en algún sistema de salud del mundo:

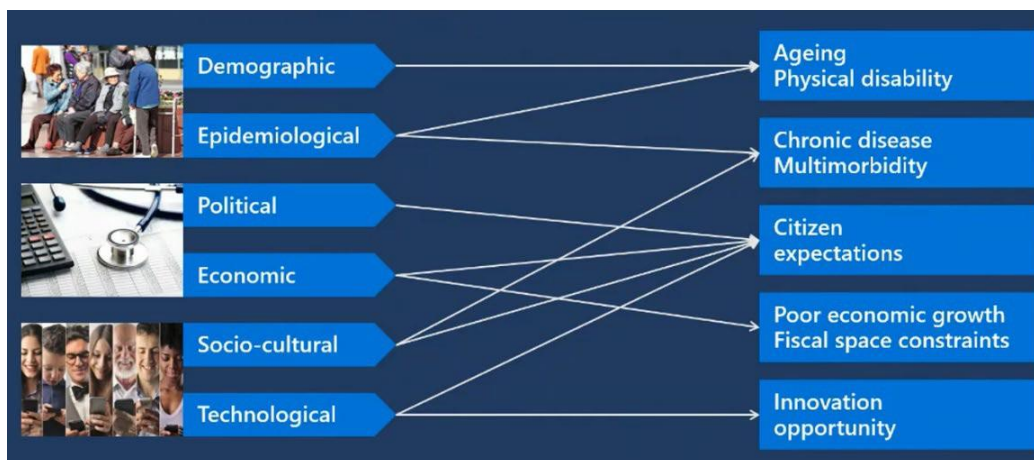
1. Factor demográfico. - Se relaciona con el envejecimiento de la población y las discapacidades físicas que esto genera de manera natural. El promedio de vida en la población, en todo el mundo se está incrementando y será una tendencia a nivel global debido a los propios avances en los modelos de salud y la calidad de vida derivada de la lucha contra la pobreza y la mejora económica. El incremento en el número de años que vive una persona trae a los sistemas de salud y de seguridad social nuevos retos por las discapacidades físicas que se irán produciendo. En un par de décadas más, muchos países europeos tendrán una población adulta que supere los 60

años y que demande servicios de salud específicos y una población de jóvenes reducida que tendrá que dar soporte a su economía.

2. Factor *ep/demiológico*. – Tiene relación con el envejecimiento de las personas su discapacidad física y las enfermedades crónicas que se irán presentando en este sector de la población. La morbilidad será múltiple y requerirá, a su vez, la asignación de recursos (humanos, materiales y financieros) específicos para darle soporte.
3. Factor político. – Está relacionado con la visión de los gobiernos y las expectativas que tienen los ciudadanos en los servicios de salud. Las prioridades de inversión de los gobiernos y las necesidades del momento a veces no son compatibles con los cambios tecnológicos y éstos, por lo general, siempre van más avanzados que los cambios políticos y las exigencias sociales.
4. Factor económico. – Se relaciona con las expectativas ciudadanas y el pobre o restringido crecimiento de la economía en varios países, así como con las limitaciones al espacio de infraestructura física que existe. Se requiere invertir en tecnología, pero que esa tecnología efectivamente sea compatible y funcione en los sistemas de salud tomando en cuenta los demás factores que están presentes.
5. Factor sociocultural. – Está relacionado con el factor de las expectativas ciudadanas y con el pobre crecimiento económico limitado de los países; pero también, con la forma en que el ciudadano entiende a los servicios de salud y el cuidado de su salud. En muchos países latinoamericanos no existe una cultura del cuidado de la salud ni hay un modelo de salud preventiva eficiente que permita prevenir enfermedades derivadas de una mala alimentación, sedentarismo, exceso de azúcares, etcétera.
6. Factor tecnológico. – Tiene que ver con las oportunidades de innovación, las cuales para los gobiernos siempre son reducidas y para las grandes empresas siempre son amplias. Esta disparidad de escenarios influye en la construcción de sociedades de la salud que utilicen la tecnología para estar interconectadas y funcionar con expedientes digitalizados y consultas en

línea, así como el uso de equipos para interpretación clínica y tratamientos basados en inteligencia artificial.

Factores existentes en los sistemas de salud



Fuente: Junaid Bajwa, “AI & Health: We rise by lifting others”, en Data Science & Law Forum 3.0⁴²⁶.

Adicionalmente, Junaid Bajwa⁴²⁷, ubica algunos indicadores que están relacionados con los factores anteriores y que también se deben de tomar en cuenta por los gobiernos, empresas y científicos para la planeación de políticas públicas o desarrollo de productos o servicios a la población mundial. Toda gira alrededor de los datos y la manera en la que se utilicen. El éxito de la atención médica digital está en el correcto almacenamiento y uso de los datos, esto hará a la tecnología más exacta, con mayores posibilidades de predicción y un nivel de sesgo menor.

⁴²⁶ Desde 2018, Microsoft organiza el *Data Science & Law Forum* como un espacio para la reflexión colectiva y el aprendizaje sobre el desarrollo de principios para una IA responsable. En 2021, las discusiones fueron de la inteligencia artificial en la práctica y la necesaria adopción de reglas que protejan los sistemas de inteligencia artificial. El material del evento puede consultarse en el sitio web:

https://datasciencelawforum.eu/?utm_campaign=833886_Thank%20you%20for%20joining%20the%20Data%20Science%20%26%20Law%20Forum%203.0&utm_medium=dotdigital&utm_source=Forum%20Europe&dm_i=4SEG%2CHVFI%2C4O3GPN%2C246OF%2C1 [15 de julio de 2022]

⁴²⁷ Junaid Bajwa, “AI & Health: We rise by lifting others”, en Data Science & Law Forum 3.0, 2021. Consúltese en:

https://datasciencelawforum.eu/?utm_campaign=833886_Thank%20you%20for%20joining%20the%20Data%20Science%20%26%20Law%20Forum%203.0&utm_medium=dotdigital&utm_source=Forum%20Europe&dm_i=4SEG%2CHVFI%2C4O3GPN%2C246OF%2C1 [15 de julio de 2022]

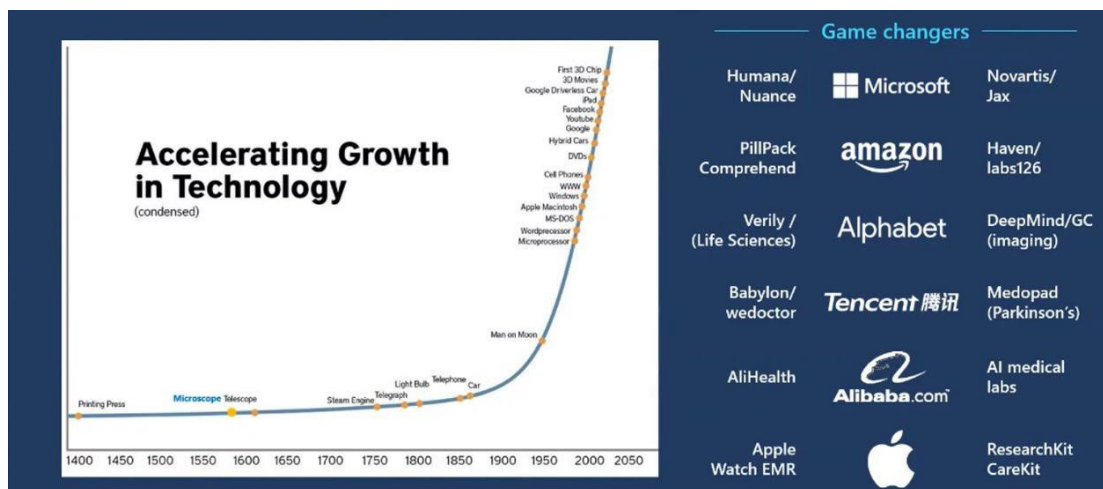
- a. Expectativas de satisfacción. - Tiene que ver con la necesidad de incrementar las expectativas de satisfacción en el uso de los servicios de salud ya que el 81% de los pacientes están insatisfechos con su experiencia de atención médica actual, que es la tradicional, la que no está utilizando la tecnología y que heredamos del siglo XIX.
- b. Falta de continuidad en el modelo de salud. - Desde antes de la llegada del COVID-19, los sistemas de salud carecen de continuidad en los tratamientos de prevención y atención de enfermedades. Se estima que, en 2020, el 70% de los procedimientos de atención médica se cancelaron, lo que provocó que los ingresos se desplomaran mientras aumentaban los gastos. La interrupción en el sistema de salud se debió al COVID-19 y a las necesidades de atención médica en rápida evolución
- c. Carencia del personal de la salud. – Existe escasez de trabajadores de la salud en todo el mundo. Se estima que para 2030, los prestadores de servicios de salud en todo el mundo tendrán un déficit de 14 millones de trabajadores. Algo que obliga a la automatización de procesos ya que las instituciones solo pueden resolver este tipo de problemas por la vía de la contratación de personal o la automatización de actividades.
- d. Sistemas fragmentados, desconectados o en silos. - Para 2022, el 50% de las grandes organizaciones del mercado de los servicios de salud se habrán fusionado para unificar los canales de participación. Esto trae ahorro de recursos y crecimiento tecnológico, pero también, puede generar una experiencia del cliente inconexa o por silos ya que los datos pueden quedar desconectados del total del sistema.
- e. Dificultad para acceder a datos de atención médica significativos. - El 41% de los proveedores dice que los desafíos de datos y análisis les impiden tener éxito en los modelos de atención basados en valores. Y esto es muy común en todo el mercado digital, no solo en los servicios de salud. Los datos es el petróleo o la energía con la que funcionan los sistemas de IA y demás herramientas tecnológicas. Sin embargo, su contenido tiene tantas deficiencias como el mismo número de usuarios del sistema de salud. Se

debe desarrollar una política global para la digitalización, almacenamiento y utilización de los datos con filtrados específicos auditables para garantizar la privacidad del paciente y la calidad del diagnóstico.

- f. Presión en seguridad y cumplimiento. - \$6.2 mil millones de dólares perdidos debido a violaciones de datos en la industria de la salud cada año. La industria de la salud lucha permanentemente por conseguir datos, pero esos datos están regulados por normas sobre privacidad que se deben respetar al momento de su uso. Las áreas de *Compliance* y la política de gobierno corporativo de las empresas de salud deben cuidar esto para evitar violaciones de datos a pacientes ya que generan pérdidas económicas serias.
- g. Ritmo de cambio tecnológico acelerado. – El cambio en la tecnología siempre ha sido mucho mayor que los cambios sociales, políticos o culturales. Hay nuevos competidores que se están fusionando con las grandes empresas y acelerando el crecimiento de la tecnología.

La tabla siguiente muestra un ejemplo de las principales empresas que están generando cambios en la industria de la salud y la tecnología:

Empresas que están compitiendo y generando cambios en la industria de la salud



Fuente: Junaid Bajwa, "AI & Health: We rise by lifting others", en Data Science & Law Forum 3.0

Si tomamos como ejemplo el área de oncología, desde la materialización de los síntomas, el diagnóstico, sus derivaciones, la experiencia del médico en tratamientos y cuidados, así como la rehabilitación y supervisión (o paliación), nos encontramos ante un panorama complejo en el que la medicina y la tecnología están buscando trabajar en conjunto para beneficio de los pacientes.

La complejidad de los tratamientos oncológicos y la personalización de los diagnósticos y tratamientos son un reto para la medicina y para la tecnología. Tan sólo pensemos lo que implica la búsqueda de información e interpretación de ésta en un mundo con altas cantidades de datos. La interpretación de estudios versus la velocidad en la que se genera la innovación en esa misma área de la medicina es incompatible para un ser humano. “Cada 30 segundos se publica un artículo médico; se suben 4,000 nuevos artículos en PubMed todos los días; 50 millones de publicaciones médicas en bases de datos públicas (no todas las publicaciones tienen necesariamente la misma calidad (sesgos inherentes); el conocimiento médico se duplica cada 73 días. Ante esto... ¿Qué pasaría si la atención médica y la informática “hablaran el mismo idioma” y esta información pudiera democratizarse y conducir a conocimientos prácticos?”⁴²⁸

Parte de la solución está en el respaldo que dan las tecnologías basadas en *Machine Learning* y en el apoyo que se está generándose desde la radiología⁴²⁹ y oncología⁴³⁰ hasta oftalmología⁴³¹ y toma de decisiones médicas⁴³². A la vez, se ha

⁴²⁸ Junaid Bajwa, “AI & Health: We rise by lifting others”, en Data Science & Law Forum 3.0. Imaginemos también el uso adecuado de wearables y apps enfocadas a los servicios de salud. El cuerpo se convierte en un área de prevención de la salud en tiempo real, que nos permite conocer y prevenir, siempre y cuando exista el consentimiento explícito del paciente, lo que le ocurre.

⁴²⁹ Wang, X. et al. Chest X-ray: hospital-scale chest X-ray database and benchmarks on weakly supervised classification and localization of common thorax diseases, en *Advances in Computer Vision and Pattern Recognition*, Springer (2019).

⁴³⁰ Patel, N. M. Enhancing Next-generation sequencing-guided cancer care through cognitive computing. *Oncologist* 23, 179–185, <https://doi.org/10.1634/theoncologist.2017-0170> (2018). [15 de julio de 2022]

⁴³¹ Brown, J. M., et al., “Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks”. *JAMA Ophthalmol*, volume 136, 2018, pp. 803-810. <https://doi.org/10.1001/jamaophthalmol.2018.1934> [15 de julio de 2022]

⁴³² Horng S., et al Creating an automated trigger for sepsis clinical decision support at emergency department triage using *Machine Learning*. *PLoS ONE* (2017). Sheth, D. & Giger, M. L. Artificial

confirmado que su uso trae riesgos inminentes relacionados con la transparencia de la información, el sesgo inherente natural de los datos con los que se alimentan y cuestiones de seguridad al usuario. Un ingrediente necesario para dar certeza o seguridad jurídica a los usuarios es la regulación de dichas tecnologías por las agencias de gobierno de cada Estado o por los organismos supranacionales.

intelligence in the interpretation of breast cancer on MRI. *J. Magn. Reson. Imaging* 51, 1310–1324, <https://doi.org/10.1002/jmri.26878> (2019). [15 de julio de 2022]

B) Watson Health: características y problemática de su uso⁴³³

Como hemos dicho, IBM crea a Watson en 2007, como el primer sistema de computación cognitiva diseñado para analizar preguntas que los usuarios pueden plantear en lenguaje natural y generar respuestas profundas con una base de evidencias previamente analizadas, mediante la comprensión del lenguaje natural, tal como lo hace un ser humano (además de aprender y mejorar sus conocimientos durante el tiempo). La computación cognitiva, es el término que se usa para describir sistemas de inteligencia artificial que tienen como objeto simular la mente humana. La cognición humana implica el análisis en tiempo real del entorno, el contexto y la intención, entre muchas otras variables que conforman la capacidad de una persona para razonar y resolver problemas.⁴³⁴

A decir de los investigadores de estos temas, se requieren varias tecnologías de inteligencia artificial para que un sistema informático construya modelos cognitivos que imiten los procesos de pensamiento humano, incluido los SED, el aprendizaje automático, el aprendizaje profundo, las redes neuronales, la PNL y el análisis de sentimientos.⁴³⁵ Unirlos todos y que funcionen similar al pensamiento humano no es fácil, pero IBM afirma que Watson lo ha logrado.

Watson Health es considerado el primer sistema de computación cognitiva enfocado al mercado de la salud. Si bien Watson, fue desarrollado específicamente para competir en el programa de televisión *Jeopardy* (2011) ganando el primer lugar, IBM decidió llevar sus capacidades cognitivas a otras actividades, entre ellas los servicios de salud. Se pensó que Watson, podía ayudar a los médicos en el proceso de toma de decisiones (diagnóstico y tratamiento). Para dimensionar este avance, en 2011 se mencionaba en los medios de comunicación que Watson estaba

⁴³³ Para más detalles y actualización de temas se puede seguir el sitio: <https://www.ibm.com/mx-es/watson-health> [15 de julio de 2022]

⁴³⁴ Cfr. Dharmendra S. Modha, Rajagopal Ananthanarayanan, Steven K. Esser, Anthony Ndirango, Anthony J. Sherbondy, and Raghavendra Singh. 2011. "Cognitive computing". *Commun. ACM* 54, 8, agosto de 2011, pp. 62 a 71. DOI: <https://doi.org/10.1145/1978542.1978559>

⁴³⁵ Bridget Botelho, "Cognitive computing", *TechTarget*, julio 2018, en <https://searchenterpriseai.techtarget.com/definition/cognitive-computing> [15 de julio de 2022]

compuesta por 90 servidores IBM Power 750 Express impulsados por procesadores de 8 núcleos, cuatro en cada máquina para un total de 32 procesadores por máquina. Los servidores funcionaban mediante una implementación de máquina virtual basada en Kernel (KVM), creando un clúster con una capacidad de procesamiento total de 80 teraflops⁴³⁶.

Watson, podía procesar hasta 500 gigas por segundo o un millón de libros por segundo, se utilizaron varios lenguajes de programación para codificar Watson (Java y C ++), que son objetos o lenguajes de programación en *Prolog*, que está asociado con inteligencia artificial. Se utilizó Linux en el sistema operativo. Los componentes de Watson incluyeron una infraestructura para el análisis de datos no estructurados de *Apache Unstructured Information Management Architecture* (UIMA). Hadoop de Apache, era un marco de programación gratuito basado en Java que admitía el procesamiento de grandes conjuntos de datos en un entorno informático distribuido *SUSE Enterprise Linux Server 11*, también tenía el sistema operativo más rápido disponible: 2,880 núcleos de procesador 15 terabytes de RAM. 500 gigabytes de información preprocesada. El software DeepQA de IBM, servía para la recuperación de información que incorpora el procesamiento del lenguaje natural (NLP) y el aprendizaje automático.⁴³⁷

La primera implementación comercial de *Watson Health* se produjo en 2013 cuando el Centro Oncológico Memorial Sloan Kettering comenzó a utilizar el sistema para recomendar opciones de tratamiento para pacientes con cáncer de pulmón basado en evidencias para garantizar que recibieran el tratamiento adecuado y reducir los costos. Cuando el personal médico ingresaba sus preguntas, Watson

⁴³⁶ Un teraflop te da un billón de operaciones por segundo.

⁴³⁷ Cfr. Lucas Mearian, Can anyone afford an IBM Watson supercomputer?, febrero 2011, <https://www.computerworld.com/article/2513312/can-anyone-afford-an-ibm-watson-supercomputer--yes-.html>

generaba una lista de hipótesis y ofrecía opciones de tratamiento para que los médicos las considerasen.⁴³⁸

Actualmente, *Watson Health* tiene las siguientes capacidades: 1) leer e intercambiar datos estructurados y no estructurados, 2) leer artículos de publicaciones especializadas en medicina, 3) analizar redes sociales y entradas de blog en varios idiomas; 4) puede generar razonamientos, 5) puede analizar la entrada de información de los pacientes para determinar, con los resultados de pruebas actualizadas, 6) las opciones de tratamiento disponibles; 7) puede examinar los síntomas del paciente, 8) buscar en todas las fuentes de datos disponibles en la red y 9) determinar hipótesis y probarlas; 10) puede ser entrenado para reconocer características asociadas con enfermedades específicas; y, 11) pueden comparar, rápidamente alguna imagen en su base de datos y proporcionar al médico un diagnóstico con el posible tratamiento.⁴³⁹

Watson Health está realizando diversas funciones en muchos hospitales del mundo. Para efectos de esta investigación hemos seleccionado ocho que están relacionadas con el tema de la responsabilidad (moral o jurídica). Aclarando nuevamente, que el estudio que se presenta toma como caso a los sistemas de IA basada en Machine Learning – Deep Learning y creados para la toma de decisiones (especialmente las del médico). Las funciones que fueron seleccionadas son las siguientes⁴⁴⁰:

1. Interpretación de imágenes médicas. - Los médicos pueden alimentar a *Watson Health* con imágenes de pacientes recién diagnosticados, que al ingresarlas se comparan con otras imágenes de su data set. Esta información

⁴³⁸ *Idem*. Consúltase también: Ed Burns, IBM Watson supercomputer, en TechTarget, junio de 2018, consúltase en: https://searchenterpriseai.techtarget.com/definition/IBM-Watson-supercomputer?_gl=1*1pwzd5r*_ga*MTM2NTMyNjI2OC4xNjIwODg1OTI5*_ga_RRBYR9CGB9*MTYyMDg4NTkyOC4xLjEuMTYyMDg4NTk0MS4w&_ga=2.267401990.963224639.1620885929-1365326268.1620885929

⁴³⁹ *Idem*.

⁴⁴⁰ Steve Smith, 5 Ways The IBM Watson Is Changing Health Care, From Diagnosing Disease To Treating It, 17 de diciembre de 2015, en Medical Daily, en <https://www.medicaldaily.com/5-ways-ibm-watson-changing-health-care-diagnosing-disease-treating-it-364394>

permite a los médicos tener una comprensión del tema, sus características y los tipos de afectaciones a la salud del paciente (presente y futuro)

2. Tratamiento de enfermedades raras. - IBM se asoció con Boston Children's Hospital para ayudar a identificar, diagnosticar y tratar enfermedades pediátricas raras. Inicialmente, la asociación se centrará en el síndrome nefrótico resistente a los esteroides (SRNS), una rara forma genética de enfermedad renal que afecta a niños de tan solo 2 años. Watson fue vinculado a la nefrología, una forma de estudio que se centra en la función renal, leyendo literatura médica sobre el tema y reuniendo más información sobre mutaciones. Watson analiza la literatura científica y las bases de datos clínicas que se encuentran en *Watson Health Cloud* con el fin de hacer coincidir mutaciones genéticas con otras enfermedades y encontrar mejores maneras para que los médicos identifiquen las opciones de tratamiento. El propósito de todo el proyecto es proporcionar a los médicos un sistema que pueda interpretar la secuencia del genoma de un niño, que, cuando se combina con la literatura médica, les permitirá identificar adecuadamente las enfermedades raras más rápido que nunca.
3. Tratamiento de pacientes con cáncer donde los médicos son escasos. - En lugares donde los pacientes superan con creces a los médicos, como la India, que apenas tiene un médico por cada 1,700 pacientes- recibir el tratamiento requerido puede ser tardado y complicado. IBM y Manipal Hospital, la tercera red de atención médica más grande de la India, se han asociado para utilizar Watson para diagnosticar y tratar el cáncer en las 16 instalaciones y centros académicos de Manipal, que tratan a unos 200,000 pacientes con cáncer cada año. Watson utiliza la información médica de revistas y libros de texto, y luego pone esta información en términos que cualquiera podría usar para encontrar diferentes opciones de tratamiento, analizar información específica del paciente y adaptar la evidencia que encuentra para satisfacer sus

necesidades de salud. También utiliza el aprendizaje profundo para actualizar continuamente sus bases de datos con más información.⁴⁴¹

4. Perfil genético de una persona y opciones de tratamiento personalizadas. - En colaboración con 14 institutos oncológicos, IBM espera reducir este tiempo de respuesta para el análisis de genética humana (a solo algunos minutos, en vez de varias horas o días). Watson está en el Centro del Genoma de Nueva York, el Centro oncológico de la Universidad de Kansas y el Centro oncológico de Yale. Estos centros utilizan las habilidades de aprendizaje profundo de Watson para traducir el ADN de un paciente en un perfil genético utilizable para varias actividades. Cuando se combina la búsqueda con literatura médica, permite a los médicos proporcionar a los pacientes las opciones de tratamiento más personalizadas.
5. Recomendaciones nutricionales. - Si bien la mayoría de las innovaciones antes mencionadas se encuentran en hospitales y centros de emergencia de todo el mundo, IBM también está llevando a Watson a los teléfonos inteligentes. IBM presentó la aplicación *Nutrino Powered* por Watson. Una colaboración con Nutrino Inc., desarrolladora de aplicaciones de alimentación saludable, la aplicación para teléfonos inteligentes está destinada a dar a las mujeres embarazadas consejos de nutrición basados en la ciencia, personalizados y contextuales. Una vez que una mujer embarazada se registra en la aplicación, puede introducir su estado de embarazo, necesidades dietéticas, objetivos de salud, preferencias alimentarias, hábitos alimenticios e incluso datos extraídos de dispositivos portátiles como Fitbit.

442

⁴⁴¹ Watson for Oncology, fue diseñado en asociación con el Memorial Sloan Kettering Cancer Center en Nueva York, considerado el centro de cáncer más grande y antiguo del mundo (fundado en 1884). Un video explicativo de Watson y el Memorial Sloan Kettering Cancer Center se puede consultar en: <https://www.mskcc.org/videos/mskcc-and-ibm-collaborate-applying-watson-technology-help-oncologists> [15 de julio de 2022]

⁴⁴² Steve Smith, *5 Ways The IBM Watson Is Changing Health Care, From Diagnosing Disease To Treating It*, diciembre 17, 2015 en *5 Ways The IBM Watson Is Changing Health Care, From Diagnosing Disease To Treating It* (medicaldaily.com) [15 de julio de 2022]

6. Asesor de salud virtual para pacientes con cáncer. - La posibilidad de que cada persona que tiene cáncer pueda encontrar información adecuada para controlar su enfermedad es un obstáculo importante. La *American Cancer Society* (ACS) con el poder de la tecnología cognitiva de IBM Watson e IBM se asociaron para crear un asesor virtual para el paciente. El asesor virtual analiza la información del sitio web de la asociación y hace recomendaciones para el paciente.
7. Soporte para la decisión clínica en cirugía robótica. - Actualmente se usa solo para pasos de rutina en procedimientos simples como cirugía ocular con láser y trasplantes de cabello. Los hospitales están introduciendo herramientas para aplicaciones como la predicción de un shock séptico, pero aún no han demostrado su valor.
8. Análisis de imágenes enfermería virtual. - Los expertos recién están comenzando a utilizar sistemas automatizados para ayudarlos a examinar radiografías, escaneos de retina y otras imágenes. Los sistemas rudimentarios pueden controlar a los pacientes entre las visitas al consultorio y proporcionar alertas automáticas a los médicos.

Para lograr todo lo anterior, IBM tuvo que crecer en infraestructura y se asoció (o adquirió) empresas dedicadas al sector de la salud y los datos. Por ejemplo, en 2015 se asoció con Apple, Johnson & Johnson y Medtronic para facilitar que las organizaciones de atención médica almacenen y analicen los datos de los pacientes utilizando las capacidades cognitivas de Watson y la nube de IBM, para proporcionar mejores conocimientos en tiempo real y recomendaciones que pueden mejorar la salud personal y la atención de enfermedades crónico-degenerativas.⁴⁴³

⁴⁴³ "IBM compra dos startups y crea *Watson Health*, unidad de negocio orientada al sector de la salud, abril 14, 2015", *Diarioti*, en <https://diarioti.com/ibm-compra-dos-startups-y-crea-watson-health-unidad-de-negocio-orientada-al-sector-de-la-salud/87178>

Adquirió también *Explorers* y obtuvo uno de los conjuntos de datos clínicos más grandes del mundo que equivale a más de 50 millones de vidas de compilación de datos de sistemas de registros médicos, operativos, financieros de más de 360 hospitales y más de 317 mil proveedores. Posteriormente, compró *Merge Health Care* y obtuvo acceso a más de 30 mil millones de imágenes médicas y 7,500 sitios de atención a la salud.

La última adquisición de IBM fue *Truven Health Care* (2016), que compró por 2.6 millones de dólares. Truven es un proveedor de análisis de datos de atención médica y conocimientos que aporta a más de 8,500 clientes a la cartera de salud de Watson y que incluye agencias gubernamentales federales y estatales de EE. UU.⁴⁴⁴

En el área genómica, ahora *Watson Health* se asoció con *Quest Diagnostics* para combinar la secuenciación genética de tumores sólidos de última generación con su poderoso análisis de datos. El médico envía una muestra para hacer una búsqueda, *Diagnostics* secuencia la estructura genómica de los tumores para ayudar a los médicos y comprender las mutaciones específicas que causan el cáncer. Luego Watson compara los resultados del paciente con los estudios clínicos de la literatura médica relevante y las reglas creadas por los principales centros de cáncer del mundo y finalmente un director médico de diagnóstico revisa los informes de Watson. El director confirma (o rechaza) los resultados y prepara un informe para cada consideración. Finalmente, el médico tratante recibe un informe del equipo médico de *Quest* y proporciona información sobre terapias y ensayos clínicos que han sido personalizados para el paciente.

En mayo de 2018, el mayor sistema de atención médica especializada de la India, Apollo, acordó adoptar Watson Oncología y Watson Genómica. Las dos plataformas de computación cognitiva de IBM ayudarán a los médicos a tomar decisiones para la atención personalizada del cáncer. El uso de Watson por parte

⁴⁴⁴ Steve Lohr, IBM Buys Truven for \$2.6 Billion, Adding to Trove of Patient Data, *The New York Times*, febrero 18, 2016, en <https://www.nytimes.com/2016/02/19/technology/ibm-buys-truven-adding-to-growing-trove-of-patient-data-at-watson-health.html>

de IBM para resolver algunos de los mayores problemas relacionados con la atención al paciente y el uso de conocimientos basados en datos para recomendar opciones de diagnósticos y tratamientos pretende demostrar el valor de Watson en la salud. En la India los médicos del *Manipal Comprehensive Cancer Center* evaluaron a Watson en 638 casos de cáncer de mama y encontraron una tasa de concordancia del 73 por ciento en comparación con las recomendaciones de tratamiento humano.⁴⁴⁵

Desde 2011 IBM ha hecho casi 50 anuncios sobre asociaciones que tenían como objetivo desarrollar nuevas herramientas para la medicina habilitadas por IA. Algunas fueron construidas para médicos e instituciones; otras se enfocaron a aplicaciones para consumidores. IBM dice que los esfuerzos de investigación han sido valiosos y que hay muchas actividades en curso. La siguiente tabla muestra el panorama de las asociaciones o alianzas con mayor precisión:

⁴⁴⁵ Diariti, "IBM compra dos startups y crea *Watson Health*, unidad de negocio orientada al sector de la salud", abril 14, 2015, *Diarioti*, en <https://diarioti.com/ibm-compra-dos-startups-y-crea-watson-health-unidad-de-negocio-orientada-al-sector-de-la-salud/87178>.

Asociaciones o alianzas para fortalecer *Watson Health*

DATE	IBM PARTNER	PROJECT	CURRENT STATUS
2011	Feb. Nuance Communications	Diagnostic tool and clinical-decision support tools	No tools in use
	Sept. WellPoint (now Anthem)	Clinical-decision support tools	No tools in use
2012	March Memorial Sloan Kettering Cancer Center	Clinical-decision support tool for cancer	Watson for Oncology
	Oct. Cleveland Clinic	Training tool for medical students; clinical-decision support tool	No tools in use
2013	Oct. MD Anderson Cancer Center	Clinical-decision support tool for cancer	No tool in use
2014	March New York Genome Center	Genomic-analysis tool for brain cancer	No tool in use
	June GenieMD	Consumer app for personalized medical advice	No app available
	Sept. Mayo Clinic	Clinical-trial matching tool	Watson for Clinical Trial Matching
2015	April Johnson & Johnson	Consumer app for pre- and postoperation coaching; consumer app for managing chronic conditions	No apps available
	April Medtronic	Consumer app for personalized diabetes management	Sugar.IQ app
	May Epic	Clinical-decision support tool	No tool in use
	May University of North Carolina, others	Genomic-analysis tool for cancer	Watson for Genomics
	July CVS Health	Care-management tool for chronic conditions	No tool in use
	Sept. Teva Pharmaceuticals	Drug-development tool; consumer app for managing chronic conditions	No tool in use; no app available
	Sept. Boston Children's Hospital	Clinical-decision support tool for rare pediatric diseases	No tool in use
	Dec. Nutrina	Consumer app for personalized nutrition advice during pregnancy	No app available
	Dec. Novo Nordisk	Consumer app for diabetes management	No app available
	2016 Jan. Under Armour	Consumer app for personalized athletic coaching	No app available
	Feb. American Heart Association	Consumer app for workplace health	No app available
	April American Cancer Society	Consumer app for personalized guidance during cancer treatment	No app available
2016	June American Diabetes Association	Consumer app for personalized diabetes management	No app available
	Oct. Quest Diagnostics	Genomic-analysis tool for cancer	Watson for Genomics from Quest Diagnostics
	Nov. Celgene Corp.	Drug-safety analysis tool	No tool in use
2017	May MAP Health Management	Relapse-prediction tool for substance abuse	No tool in use

Fuente: Eliza Strickland, *IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care* <https://ieeexplore.ieee.org/document/8678513>

Diversos analistas han cuestionado a IBM por dedicarse a este tipo de inversiones sin tener con precisión el panorama de las decisiones médicas, o por lo menos, sin entender si las capacidades del sistema *Watson Health* pueden llegar a producir el conocimiento del médico, el cual, como hemos visto tiene varios ingredientes poco controlables, como por ejemplo el conocimiento heurístico y el grado de exactitud en el diagnóstico y tratamiento de las enfermedades como el cáncer.

Creemos que Watson tiene un serio obstáculo para un avance mayor y tiene que ver con el razonamiento heurístico del médico humano, así como los errores o sesgos que de manera natural se generan de forma que los sistemas de IA hoy por hoy, en el campo de la medicina, se ven limitados al acompañamiento médico humano, siendo muy arriesgado dejar la última decisión en la máquina.

Ante este tipo de crítica y las altas cantidades de dinero invertido, IBM se ha mantenido en una postura muy cuidadosa y, sobre todo, silenciosa. A continuación, expondremos información que revela su preocupación o, por lo menos, que demuestran que la tecnología de *Watson Health* debe ser cuidadosamente valorada si se decide continuar con su uso en el área médica.

En las demostraciones que se han hecho de *Watson Health*, no se pone en duda su capacidad intelectual y su potencial para la búsqueda de información de todo tipo (hasta síntomas de pacientes con enfermedades raras). En general, sus diagnósticos y tratamientos recomendados tienen un nivel de confianza aceptable, sobre todo por el soporte que le dan los diversos enlaces a literatura médica con soporte científico. Sin embargo, no todas las implementaciones requieren únicamente de la capacidad intelectual de *Watson Health*. Así como es de humanos errar; Watson erra y tiene una problemática generada por sus propios creadores. De 2013 a la fecha, se tienen documentados varios casos que han sido cuestionados por diversos analistas y expertos.

IBM ha buscado participar en los departamentos de oncología de muchos hospitales del mundo y esto ahora le trae sorpresas en la implementación de

Watson Health. Podemos decir que Watson sirve como acompañamiento del experto médico, pero no los sustituye aún; no por el momento. Aunque no se descarta que de pronto llegue una sorpresa y esta tecnología se use como hoy se usa una prueba para diagnosticar alguna enfermedad crónico-degenerativa.

En un informe denominado Inteligencia artificial e innovación en el sector de salud pública, IBM reconoce que “Watson for Oncology aún no está totalmente desarrollado, y pueden producirse errores” y señala también que “Watson for Oncology no está capacitado para resolver tratamiento complejo por un período prolongado”.⁴⁴⁶ En un artículo publicado en el sitio STAT, que es consultado por médicos en muchas partes del mundo, narra la situación de Watson y los comentarios de médicos en 50 hospitales del mundo (13 países), quienes manifiestan que el programa Watson Oncology es una buena herramienta, pero está en su etapa de infancia.

¿Qué fue lo que pasó con IBM? ¿Por qué se buscó impactar en el mercado de los servicios de salud sin tener lo suficientemente avanzado el programa Watson? Creemos que aquí, como en otros casos de empresas exitosas, se decidió impulsar un proyecto novedoso sin medir las consecuencias o limitantes tanto del médico como de la empresa y se crearon expectativas altas. La empresa, como decía Carlos Llano Cifuentes, es “una comunidad de personas, que un entramado de relaciones puramente mercantiles, en donde incluso se crea un ambiente inhóspito”⁴⁴⁷ y su convivencia cotidiana acarrea responsabilidades. En este sentido, los directivos de IBM debieron crear pautas de actuación ética y legal, utilizar un modelo GRL (*Governance, Risk and Compliance*) o un *Compliance and Ethics Management System* así como vincular al área de Compliance para hacer una medición de riesgos legales, éticos y, principalmente, tecnológicos, para “arrancar

⁴⁴⁶ Suspenden a IBM Watson por recomendar “tratamiento mortal” para pacientes de cáncer, Diarioti, noviembre 1, 2017, consúltase en: <https://diarioti.com/suspenden-a-ibm-watson-por-recomendar-tratamiento-mortal-para-pacientes-de-cancer/106013> [15 de julio de 2022]

⁴⁴⁷ Carlos Llano sobre el libro “Dilemas éticos de la empresa contemporánea”, en el sitio: https://www.youtube.com/watch?v=i5W_8_pSzsQ [20 de julio de 2022]

los conceptos de la realidad; en lugar de aplicarlos rígida y homogéneamente a ella”.⁴⁴⁸

IBM seguramente estaba pensando en la *responsabilidad del consecuente* (es decir, la responsabilidad de las consecuencias de sus acciones y como sus acciones no estaban fuera de la norma, porque no había norma, no midió otros riesgos). Si hubiera aplicado la que denominamos: *Teoría de una Moral de la Responsabilidad Empresarial*, de Carlos Llano Cifuentes no estaría pensando en la venta de *Watson Health*.

El 18 de febrero de 2021, sorprendió la noticia que apareció en varios medios de comunicación en la que se decía que IBM *Watson Health* sería vendido a finales de 2021.⁴⁴⁹ La nota de prensa explica que IBM está estudiando alternativas para la unidad que podrían incluir una venta⁴⁵⁰ a una firma de capital privado o un actor de la industria o una fusión con una compañía de cheques en blanco, dijeron las personas. Y concluye: “La unidad, que emplea inteligencia artificial para ayudar a hospitales, aseguradoras y farmacéuticas a administrar sus datos, tiene aproximadamente mil millones de dólares en ingresos anuales y actualmente no es rentable, dijeron las personas.” La rentabilidad no se produjo o quizá no se produjo como se había esperado.⁴⁵¹

Una investigación de STAT descubrió que su caída se debió a una serie de heridas autoinfligidas que se acumularon durante años de marketing sobrecalentado y productos de bajo rendimiento. Los documentos internos y las entrevistas con exempleados (que renunciaron o fueron removidos por este tema) y colaboradores de la industria apuntan a un liderazgo corporativo que priorizó la publicidad y los objetivos financieros a corto plazo sobre el trabajo lento de la ciencia

⁴⁴⁸ Carlos Llano Cifuentes, Presentación del libro: *La enseñanza de la dirección y el método del caso*, en el sitio del IPADE: <https://www.carlosllanocatedra.org/la-ense%C3%B1anza-de-la-direcci%C3%B3n-y-el-m%C3%A9todo-del-caso-rese%C3%B1a> [15 de julio de 2022]

⁴⁴⁹ Laura Cooper y Cara Lombardo, “IBM Explores Sale of IBM *Watson Health*”, *The Wall Street Journal*, febrero 18, 2021, en <https://www.wsj.com/articles/ibm-explores-sale-of-ibm-watson-health-11613696770>

⁴⁵⁰ <https://medika.life/fifty-of-the-most-impactful-tech-innovators-in-digital-healthcare-in-2021/> [15 de julio de 2022]

⁴⁵¹ *Idem*.

y la construcción del tipo de negocio que podría sobrevivir a su exageración inicial. Este tema cae en las responsabilidades de la empresa de las que habla Carlos Llano y de su teoría que explicaremos a mayor detalle en una sección posterior. Por lo pronto diremos que “las responsabilidades que debió considerar, además de la responsabilidad del consecuente, son: la responsabilidad del antecedente (o responsabilidad de los principios), la responsabilidad congruente (o responsabilidad de los proyectos anteriores de mi vida); y la responsabilidad trascendente (o de la misión a la que, con mi vida, he sido destinado)”.

Algunos oncólogos dicen que confían más en su propio juicio que en Watson. Es decir, el ser humano prefiere al ser humano, aunque la máquina tenga probabilidades estadísticas de mayores aciertos. Otros dicen que Watson sugiere solo tratamientos estándar que conoce bien y no puede intervenir en temas desconocidos. Y también, algunos médicos lo encuentran útil como una segunda opinión instantánea que pueden compartir con los pacientes nerviosos.

Estos comentarios aunados a la fuerte campaña de ventas de IBM han influido en su permanencia dentro del mercado de los servicios de salud. Al grado que *Watson Health* ha tenido más suerte fuera de los Estados Unidos, con hospitales en India, Corea del Sur, Tailandia y más allá adoptando la tecnología. Muchos de estos hospitales utilizan con orgullo la marca IBM Watson en su marketing, y les dicen a los pacientes que recibirán atención oncológica impulsada por la inteligencia artificial. Sin embargo, IBM no ha podido incrementar este negocio hacia los porcentajes de ventas estimadas.

Otro tema que genera un riesgo inevitable tiene que ver con lo legal: mucho de lo que Watson hace no está regulado con la precisión debida. Las autoridades y los particulares se mantienen sin una supervisión de los modelos de IA como Watson Health y no se va más allá del contrato privado entre las empresas que prestan servicios de salud y la creadora (o administradora) del sistema (IBM) y no dimensionan la importancia de intervenir o por lo menos de comenzar a analizar tres supuestos que se presentan en la práctica. El primero, si las categorías actuales de la responsabilidad jurídica sirven a la empresa para solucionar cualquier

problemática generada por un mal uso de o mal funcionamiento de un sistema automatizado. Si la respuesta es negativa, entonces estamos ante un segundo supuesto: ¿qué se debe hacer para proteger a la empresa?; es decir, si a la par de una respuesta jurídica debiera haber una respuesta ética dentro de la empresa (derivada de su propio gobierno corporativo). Evidentemente, la gestión de riesgo y el *Compliance* son las herramientas más idóneas para proteger a la empresa como parte de las medidas de autorregulación que el *soft law* ofrece. Y la tercera pregunta, en esta dinámica reflexiva es: Ante una falla o mal funcionamiento del sistema y sin posibilidad de utilizar las categorías jurídicas de la responsabilidad ¿quién es el responsable por el daño causado ante un mal uso o funcionamiento erróneo del sistema automatizado? ¿La agencia humana debe asumir toda la responsabilidad? O ¿existe una posibilidad de invocar alguna excluyente de esa responsabilidad ante la imposibilidad de conocer las razones por las cuales el sistema tomó esa decisión ya que le restó capacidades de decisión y análisis?

En estricto sentido, *Watson Health* no tiene ningún soporte regulatorio más allá de las normas de la legislación mercantil, comercial o de salud da a los instrumentos médicos, las cuales pueden verlo como un servicio, un producto de software o un instrumento para la empresa. Así como el médico utiliza el bisturí y las radiografías, la consulta a *Watson Health* puede considerarse como otra herramienta para el diagnóstico y tratamiento médicos.⁴⁵²

En México, por ejemplo, se ha venido discutiendo desde 2019 la creación de la Norma Oficial Mexicana para regular las buenas prácticas de fabricación de dispositivos médicos. La norma define al dispositivo médico y al “software como dispositivo médico” así como otros conceptos relacionados de la siguiente manera:

“instrumento, aparato, utensilio, máquina, **incluido el software** para su funcionamiento, producto o material implantable, agente de diagnóstico, material, sustancia o producto similar, para ser empleado, solo o en combinación, directa o indirectamente en seres humanos; con alguna(s) de las siguientes finalidades de uso:

⁴⁵² Casey Ross y Ike Swetlitz, “IBM to Congress: Watson will transform health care, so keep your hands off our supercomputer”, *STAT*, octubre 4, 2017, en <https://www.statnews.com/2017/10/04/ibm-watson-regulation-fda-congress/> [15 de julio de 2022]

· **Diagnóstico, prevención, vigilancia o monitoreo, y/o auxiliar en el tratamiento de enfermedades;**

- Diagnóstico, vigilancia o monitoreo, tratamiento, protección, absorción, drenaje, o auxiliar en la cicatrización de una lesión;
- Sustitución, modificación o apoyo de la anatomía o de un proceso fisiológico;
- Soporte de vida;
- Control de la concepción;
- Desinfección de Dispositivos Médicos;
- Sustancias desinfectantes;
- Provisión de información mediante un examen *in vitro* de Muestras extraídas del cuerpo humano, con fines diagnósticos;
- Dispositivos que incorporan tejidos de origen animal y/o humano, y/o
- Dispositivos empleados en fertilización *in vitro* y tecnologías de reproducción asistida;

Su finalidad de uso principal no es a través de mecanismos farmacológicos, inmunológicos o metabólicos, sin embargo, pueden ser asistidos por estos medios para lograr su función. Los Dispositivos Médicos incluyen a los Insumos para la salud de las siguientes categorías: equipo médico, prótesis, órtesis, ayudas funcionales, **agentes de diagnóstico**, Insumos de uso odontológico, materiales quirúrgicos, de curación y productos higiénicos.

Documento Maestro, al documento autorizado que contienen la información para realizar y controlar las operaciones de los procesos y actividades relacionadas con la fabricación de un producto.

[...]

Gestión de Riesgos de Calidad, al proceso sistemático para la valoración, control, comunicación y revisión de los riesgos a la Calidad de los Dispositivos Médicos a través de su Ciclo de Vida.

[...]

Rastreabilidad o Trazabilidad, a la capacidad de reconstruir la historia, localización de un elemento, un componente o de una actividad, utilizando los Registros como evidencia.

Sistema Computarizado/Computacional, a cualquier equipo, proceso u operación que tenga acoplada una o más computadoras y un software asociado o un grupo de componentes de hardware diseñado y ensamblado para realizar un grupo específico de funciones.

Software como Dispositivo Médico, al utilizado con uno o más propósitos médicos, que tiene como característica principal que no requiere de un hardware para cumplir con el propósito médico previsto; es capaz de funcionar en plataformas computacionales generales y puede utilizarse solo y/o en combinación con otros productos (ej., como módulo, otros Dispositivos Médicos, etcétera). Las aplicaciones móviles que cumplen con esta definición son consideradas Software como Dispositivo Médico. El software que hace funcionar al Dispositivo Médico físico queda excluido de esta definición.

Validación de Software como Dispositivo Médico, a la evidencia documental generada a través de la recopilación y evaluación de la capacidad de un Software para generar con exactitud, integridad y precisión las funciones previstas a partir de los datos de entrada.

5. Clasificación de los Dispositivos Médicos

5.1 Los Dispositivos Médicos se clasifican, de acuerdo con el riesgo que representa su uso, de la siguiente

manera:

5.1.1 Clase I: Dispositivos Médicos conocidos en la práctica médica cuya Seguridad y eficacia están comprobadas y que generalmente no se introducen en el organismo.

5.1.2 Clase II: Dispositivos Médicos conocidos en la práctica médica y que pueden tener variaciones en el material con el que están elaborados o en su concentración y, generalmente, se introducen al organismo permaneciendo menos de treinta días.

5.1.3 Clase III: Dispositivos Médicos nuevos o recientemente aceptados en la práctica médica, o bien que se introducen al organismo y permanecen en él, por más de treinta días.

17. Actividades subcontratadas

Fundamentos.

Cualquier actividad incluida en esta Norma que sea subcontratada debe estar definida, acordada y controlada para evitar imprecisiones que puedan dar como resultado un producto u operación de Calidad insatisfactoria.

Se debe formalizar en un contrato por escrito entre el agente contratante y el agente contratado que establezca claramente las responsabilidades de cada parte.

El Sistema de Gestión de Calidad del agente contratante debe reflejar claramente el modo en el que el responsable sanitario, o persona que autoriza la Liberación de cada Lote de producto, contempla las actividades subcontratadas en su responsabilidad.

17.4.2 El **contrato debe describir** claramente quién asume la responsabilidad de cada etapa de la actividad subcontratada, tal como: gestión de conocimiento, transferencia tecnológica, cadena de suministro, subcontratación, Calidad y adquisición de materiales, análisis y liberación de materiales, responsabilidad de la Producción y controles de Calidad (incluyendo Controles en Proceso, muestreo y análisis).⁴⁵³

De lo anterior, surge la pregunta siguiente: ¿Watson Health pudiera equiparse a un dispositivo médico o a un software como dispositivo médico? Sostenemos que no. Watson Health, al igual que otros sistemas que utilizan cómputo cognitivo o ML –

⁴⁵³ Proyecto de Norma Oficial Mexicana PROY-NOM-241-SSA1-2018, Buenas prácticas de fabricación de dispositivos médicos, DOF, 30 de octubre de 2012. Consúltase en: https://www.dof.gob.mx/nota_detalle.php?codigo=5562796&fecha=14/06/2019#gsc.tab=0 [20 de julio de 2022]

DL no puede ser clasificado como sistema de software tradicional y sería un error grave considerarlo así ya que en la práctica funciona como un sistema interactivo que lleva el control total del diagnóstico, dejando al ser humano una parte mínima, además de que está diseñado para aprender del entorno o de los datos y tomar decisiones sin intervención del ser humano.

En estricto sentido, esta tecnología supera por mucho los conceptos vertidos en el proyecto de NOM y en la posibilidad de regulación en un contrato privado. Ahora pensemos el contenido del proyecto de NOM tomando en cuenta los siguientes casos:

- a. Caso Hospital Nacional de Dinamarca. - El Hospital decide suspender el uso de Watson Oncology, luego de que propusiera utilizar un medicamento que habría tenido consecuencias fatales para los pacientes. Los médicos responsables concluyeron que Watson Oncology es una tecnología demasiado rudimentaria para el fin que quiere dársele. Leif Panduro Jensen, director del Departamento de Oncología del Hospital Nacional de Dinamarca, comentó que, en algunos casos, Watson propuso utilizar medicamentos que podrían haber causado el fallecimiento del paciente. Por ejemplo, hemos visto que Watson ha propuesto un tratamiento con sustancias que el paciente nunca más en su vida debía tomar, debido al riesgo vital que implican. Panduro explicó que para los ensayos se pidió a Watson proponer tratamientos para tres tipos de cáncer: tórax, pulmones y colon. Los médicos contrastaron las respuestas con la situación de 31 pacientes que ya habían recibido tratamiento. En dos terceras partes, los médicos concordaron con el tratamiento recomendado por Watson. Sin embargo, en la tercera parte de los casos el tratamiento propuesto por Watson era descabellado, en palabras de Panduro.⁴⁵⁴
- b. Caso MD Anderson Cancer Center of Houston. - El MD Anderson Cáncer Center en Houston lanzó un proyecto en 2013 para construir un sistema de

⁴⁵⁴ "Suspenden a IBM Watson por recomendar "tratamiento mortal" para pacientes de cáncer", *Diarioti*, noviembre 1, 2017, consúltese en: <https://diarioti.com/suspenden-a-ibm-watson-por-recomendar-tratamiento-mortal-para-pacientes-de-cancer/106013> [15 de julio de 2022]

apoyo a las decisiones impulsado por la tecnología Watson para ayudar a los médicos a determinar las mejores opciones de tratamiento. Sin embargo, se generaron reportes que, después de gastar más de \$62 millones de dólares en el proyecto (durante cuatro años), los administradores del hospital cancelaron el proyecto, diciendo que no habían logrado los objetivos por errores en sus pronósticos.⁴⁵⁵

- c. Centro Médico Gil de la Universidad Gachon, en Corea del Sur. - Sus principales recomendaciones para 656 pacientes con cáncer de colon coincidieron con las de los expertos solo el 49 por ciento de las veces. Los médicos informaron que a Watson le fue mal con los pacientes mayores, no sugirió ciertos medicamentos estándar y tenía un error que hizo que recomendara vigilancia en lugar de un tratamiento agresivo para ciertos pacientes con cáncer metastásico. Estos estudios tenían como objetivo determinar si la tecnología de Watson for Oncology funciona como se esperaba. Pero ningún estudio ha demostrado todavía que beneficie a los pacientes.⁴⁵⁶
- d. Hospital universitario de Giessen y Marburg. - Se tuvo que suspender su uso cuando un médico le dijo a Watson que un paciente sufría de dolor en el pecho y el sistema no consideró que pudiera estar sufriendo un ataque cardíaco. En cambio, sugirió una rara enfermedad infecciosa.⁴⁵⁷ En otra ocasión, Watson sugirió que se administrara un medicamento a un paciente con cáncer con sangrado severo que podría haber empeorado el sangrado.⁴⁵⁸

⁴⁵⁵Ed Burns, "IBM Watson supercomputer", en *TechTarget*, junio de 2018, consúltese en: https://searchenterprisetarget.com/definition/IBM-Watson-supercomputer?_gl=1*1pwzd5r*_ga*MTM2NTMyNjI2OC4xNjIwODg1OTI5*_ga_RRBYR9CGB9*MTYyMDg4NTkyOC4xLjEuMTYyMDg4NTk0MS4w&_ga=2.267401990.963224639.1620885929-1365326268.1620885929 [15 de julio de 2022]

⁴⁵⁶ Casey Ross y Ike Swetlitz, "IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show", *STAT*, julio 25, 2018, en <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/>

⁴⁵⁷ Martin U. Müller, "Medical Applications Expose Current Limits of AI", *Spiegel*, 3, agosto 2018.

⁴⁵⁸ Chen, Angela, "IBM's Watson Gave Unsafe Recommendations For Treating Cancer", *Verge*, 26 de julio de 2018. Carissa, Véliz, *Privacy is Power*, Transworld. Edición de Kindle, p. 239.

¿Qué pasaría si en estos casos el agente humano no toma el control de la tecnología y por alguna razón decide seguir adelante y confiar en el sistema y se genera un daño? ¿Es responsable el médico que confió en la tecnología y decidió desplazar su propio juicio por el razonamiento de Watson Health? Si ubicamos a Watson Health como un “dispositivo médico o software como dispositivo médico” y lo certificamos bajo la NOM mexicana, la responsabilidad directa y total queda en manos del ser humano más cercano al uso de esta tecnología. Esta decisión protege al sistema, pasando la totalidad de la responsabilidad del daño que se produzca al ser humano y no habría posibilidades de reprochar su responsabilidad ante una falla o mal funcionamiento del sistema. ¿No es esto un paralogismo argumentativo que lleva la responsabilidad de la máquina al ser humano? Creemos que sí, que los hombres deben ser menos reprochables de su conducta ante el derecho cuando se usan sistemas de inteligencia artificial basados en cómputo cognitivo o ML – DL. Al mismo tiempo, la máquina debería ser más responsable si se detectan sesgos algorítmicos que no son predecibles por el ser humano y que pueden ser considerados dentro de una nueva categoría de responsabilidad denominada “responsabilidad algorítmica”.

En síntesis, Watson Health no puede ser clasificado como un dispositivo médico tradicional o como una pieza de *software* controlada por la voluntad del ser humano ya que funciona como un sistema interactivo que toma decisiones sin intervención del ser humano y supera los modelos de software que han sido certificados por distintas normas. Esto se ha querido hacer con otros sistemas similares en industria tan sofisticadas como la aviación, certificando a los pilotos automáticos o directores de vuelo como una pieza más de software, pero en la práctica no funciona así.

Madeleine Clare Elish, da el soporte a nuestro argumento, al estudiar diversos casos en los que el ser humano utiliza este tipo de tecnología y entra en una zona de “deformación moral” y de afectación a su capacidad de razonamiento y decisión, atribuyéndole una responsabilidad por una acción en la que no participó ni tenía posibilidades de tener el control de la decisión que tomó del sistema

autónomo. Esto va más allá de entender a esta tecnología como parte de una categoría tradicional de los dispositivos médicos o software ya que sus implicaciones, están afectando a diversas industrias, a los sistemas de justicia y al ser humano o a sus familias que fueron víctimas de un daño causado por esta tecnología.

Al analizar varios accidentes que involucran el uso de sistemas automatizados, Clare Elish, llega a la conclusión que se forma una “zona de deformación moral” en la que la responsabilidad de una acción puede atribuirse erróneamente a un actor humano que tenía un control limitado sobre el comportamiento de un sistema automatizado o autónomo: “el ser humano en un sistema altamente complejo y automatizado puede convertirse simplemente en un componente, accidental o intencionalmente, que soporta la peor parte de las responsabilidades morales y legales cuando el sistema en general funciona mal... El concepto es a la vez un desafío y una oportunidad para el diseño y la regulación de sistemas humano-robot.”.⁴⁵⁹

⁴⁵⁹ Clare Elish, Madeleine, “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction”, *Engaging Science, Technology, and Society*, volumen 5, 2019, pp. 41 y ss. Consúltese en: DOI:10.17351/ests2019.260. La autora a través del análisis de casos que tuvieron gran impacto va demostrando cómo el ser humano se está volviendo el responsable del uso de esta tecnología independientemente de no haber participado con su voluntad en la toma de decisiones. También demuestra cómo hay una tendencia de culpar al ser humano de actos o daños producidos por la máquina al no haber posibilidad de entender su funcionamiento o el origen de la falla. El accidente del vuelo 447 de Air France, el 1 de junio de 2009, en el que fallecieron en el Océano Atlántico 228 personas, es un caso paradigmático al demostrarse que la tripulación “perdió la conciencia sobre lo que estaba pasando” y no respondieron de manera efectiva ante los problemas con los sensores de velocidad. El director de vuelo o piloto automático se desconectó y los pilotos humanos no supieron reaccionar ante lo sucedido. Adicionalmente, los sensores de velocidad dieron lecturas no válidas. El uso de sistemas de inteligencia artificial en los aviones ha mejorado la seguridad, pero disminuido la capacidad de los pilotos para entender señales y reaccionar a ellas. El juicio duró 10 años, se exculpó a Air France y los jueces determinaron que el siniestro fue responsabilidad de los pilotos que perdieron el control de la aeronave. La Fiscalía no pudo mantener su hipótesis respecto a que la compañía aérea había tenido negligencia al no capacitar o no informar a sus pilotos sobre el control de los sensores cuando el director de vuelo tuviera alguna falla. Un piloto actualmente tiene el control del avión 4 minutos, 2 al despegue y 2 en el aterrizaje, el director de vuelo controla el resto del tiempo el avión. Es decir, si el sistema autónomo se presenta como más capaz para controlar el avión, y la mayor cantidad de tiempo del vuelo es controlado por sistema de piloto automático reduciendo de tiempo controlado directamente por el piloto humano ¿quién tiene la responsabilidad jurídica del control de la aeronave?. Madeleine lanza una provocación para repensar cómo, por qué y con qué implicaciones se asignará la responsabilidad cuando fallan los sistemas automatizados, autónomos. Existe un video que narra lo ocurrido en: <https://www.youtube.com/watch?v=WC6EGAmO9Wo>

Si bien actualmente esa tecnología no ha cumplido o superado las expectativas para las que fue construida y, como dice Eric Topol, los problemas que IBM Watson encontró con el cáncer son representativos de las limitantes que tiene esta tecnología en el sector salud, sus esfuerzos por mejorar el diagnóstico en la medicina van a continuar. Ciertamente, dice Topol, existe la posibilidad de que la informática marque una gran diferencia, pero hasta ahora ha habido un cumplimiento mínimo de la promesa. La dificultad de recopilar y agregar los datos ha sido subestimada, no solo por Watson, sino por todas las empresas de tecnología que se involucran con la atención médica.⁴⁶⁰

Expertos en estos temas ubican actualmente a *Watson Health* en la tercera fase del ciclo de vida de la tecnología (o época de desilusiones); la cual, en teoría, seguiría con una pendiente de inclinación hacia los futuros ajustes y modificaciones que se realicen y con los cuales comenzaría a cristalizar frutos más prometedores (con productos modificados y actualizados de segunda o tercera generación) para llegar a la productividad plena por la adopción generalizadas de su uso. La siguiente tabla muestra el ciclo de vida de la tecnología:

⁴⁶⁰ Topol, Eric. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books. Edición de Kindle, p. 56.

Digital Health Hype Cycle 2020



Fuente: <https://www.healthcare.digital/single-post/2020/01/29/Digital-Health-Hype-Cycle-2020>

Es un hecho que el fracaso o no de Watson, no detiene a la *Healthcare Technology*. Su amplia aplicabilidad y relevancia supera hoy por hoy a IBM. Lo que nadie podrá cuestionar es que IBM fue el pionero en llevar la tecnología al diagnóstico y tratamiento del cáncer⁴⁶¹ y todos los competidores están luchando por superarlo y tener el control de ese mercado con medidas de procesamiento de datos más seguras, predecibles, eficaces y ofreciendo mayores beneficios a las personas.

El fracaso o no de IBM no detiene a su vez a otras empresas que están invirtiendo millones de dólares para aplicar la computación cognitiva en el cuidado de la salud. Google, por ejemplo, duplicó un acuerdo de datos con la cadena de hospitales Ascension.⁴⁶² y ha intensificado una asociación con *Mayo Clinic*, uno de

⁴⁶¹ Cfr. Medika Life, "Fifty of the Most Impactful Tech Innovators in Digital Healthcare in 2021", *The Medika Life Top 50 Watchlist for digital health companies impacting the healthcare industry in 2021* Fifty of the Most Impactful Tech Innovators in Digital Healthcare in 2021 | Medika Life

⁴⁶² Erin Brodwin, "Google expands controversial pilot project using patient data", *STAT*, 23 de febrero de 2021, en: <https://www.statnews.com/2021/02/23/google-project-nightingale-care-studio/>

los primeros colaboradores de IBM. Amazon y Microsoft también están expandiendo sus negocios de atención médica, ofreciendo capacidades de almacenamiento en la nube específicas de la industria y herramientas de inteligencia artificial para ayudar a los hospitales a analizar sus datos. Apple está invirtiendo en *Health Care Technology* con varias apps.

En México, desde 2016, se llevó a cabo la presentación del Consejo Consultivo del Consorcio de Cómputo Cognitivo en Salud de IBM, como parte del proyecto global de IBM *Watson Health*.⁴⁶³ En 2017, el Centro de Tecnología Cognitiva del Hospital Ángeles integró el servicio de Watson al tratamiento de personas con cáncer. La computadora más inteligente del mundo -dice el video de divulgación⁴⁶⁴- se integra al servicio del médico oncólogo y permitirá brindar opciones de tratamiento personalizadas a sus pacientes con cáncer⁴⁶⁵. Este sistema fue entrenado por especialistas del *Memorial Sloan Kettering Cancer Center* y ofrece las mejores opciones de tratamiento en oncología, busca y obtiene en minutos respuestas precisas, a través de comparar millones de datos contenidos en cientos de libros de medicina, miles de publicaciones científicas, casos clínicos e incluso notas médicas de reconocidos especialistas en todo el mundo... compara el caso también con otros casos clínicos de pacientes del *Memorial Sloan Kettering Cancer Center*, por lo que las opciones sugeridas serán siempre basadas en evidencia. Y también anuncia: “Será como tener un colega capaz e inteligente que pueda revisar toda la información reciente relacionada con mi paciente.... es veloz, riguroso, y tiene la misteriosa habilidad de entender la manera en que la evidencia disponible se aplica específicamente al paciente a quien estoy tratando.”⁴⁶⁶

⁴⁶³ Consúltase la nota: “IBM y Fundación Carlos Slim promueven en México el uso de tecnologías para la salud” en <https://fundacioncarlosslim.org/ibm-fundacion-carlos-slim-promueven-en-mexico-uso-tecnologias-la-salud/>

⁴⁶⁴ Puede consultarse el video que explica IBM *Watson Health* en el sitio: <https://www.youtube.com/watch?v=SUR8K-3x7M0&t=94s>

⁴⁶⁵ Consúltase el artículo: “Centro de Tecnología Cognitiva de Hospitales Ángeles” <https://www.hospitalesangeles.com/saludyvida/articulo.php?id=5763#>

⁴⁶⁶ Consúltase también la página: ¿Qué es IBM Watson for Oncology? <https://centrodetecnologiacognitivahospitalesangeles.com/queeswatson.php>

El 21 de enero de 2022 fue anunciada la venta de los activos de *Watson Health* a *Francisco Partners*, una empresa de San Francisco. IBM gastó miles de millones de dólares en la adquisición de empresas de información sanitaria para utilizarla en hospitales. Truven, Merge Healthcare, Explorys y Phytel, llegaron con una impresionante variedad de datos de pacientes. IBM luchó por integrar esos datos en el negocio. Sin embargo, no logró sus objetivos. Entre los mayores fracasos de IBM se encuentra *Watson for Oncology*, un producto que creó con el Centro de Cáncer Memorial Sloan Kettering para usar IA para brindar consejos sobre el tratamiento del cáncer a pacientes de todo el mundo. IBM comercializó el producto como una revolución en el tratamiento de precisión que podría ayudar a curar a más pacientes. Pero sus capacidades técnicas eran mucho más limitadas de lo que sugería el marketing y no pudo producir evidencia clínica para respaldar sus afirmaciones descomunales. Veremos más adelante quién toma el liderazgo de este tipo de servicios de salud.⁴⁶⁷

Antes de concluir este apartado, queremos reflexionar sobre el derecho humano a la salud y la importancia de fomentar el uso de la *Healthcare Technology* en un momento histórico en el que los seres humanos fuimos afectados por la pandemia del COVID-19⁴⁶⁸. Esta externalidad afectó los sistemas de salud, educación, trabajo, comercio; generó la más primera crisis económica del siglo XXI y puso al descubierto las debilidades del ser humano ante el aislamiento. No es casualidad que estos sectores, tan importantes para la vida de la humanidad, se hayan frenado o, en muchos casos, colapsado por completo. Todos mostraron las flaquezas de sus orígenes. Ni siquiera los gobiernos más poderosos o las sociedades más avanzadas lograron impedir que avanzara la pandemia. Creemos que el uso de la tecnología se ha convertido en un deber moral de todo ser humano.

La Organización Mundial de la Salud, por años ha pedido a los gobiernos que utilicen la tecnología para la modernización de sus sistemas de registro y consulta,

⁴⁶⁷ Cfr. Casey Ross, "Once billed as a revolution in medicine, IBM's *Watson Health* is sold off in parts", *STAT*, 21 de febrero de 2022, consúltese en: <https://www.statnews.com/2022/01/21/ibm-watson-health-sale-equity/>

⁴⁶⁸ De enero de 2020 a abril de 2021 se tenían contabilizadas más de tres millones de muertes en todo el mundo. Para conocer las cifras actualizadas puede acudir el sitio: <https://covid19.who.int/>

digitalizando documentos y los llamados expedientes electrónicos. Lamentablemente, el COVID-19 tomó por sorpresa a los sistemas de salud del mundo. Debemos reconocer que en gran parte de América Latina los servicios de salud -públicos, principalmente, se mantienen con debilidades serias que se deben corregir. Hoy se trabaja con un modelo de emergencia COVID-19 que no soluciona el problema ni da las bases para la modernización, pero si nos muestra el reto tecnológico en este sector.

El uso de la tecnología en los servicios de salud requiere la participación conjunta de los gobiernos, las empresas y los científicos. Ninguno puede lograr algo sin el apoyo de los otros. En 2007, la Organización Mundial de la Salud trajo a debate la importancia de que los países utilicen la tecnología para beneficio de pacientes, personal de la salud, empresas y los propios gobiernos. Y es que la ciencia ha dado grandes aportaciones durante los últimos años gracias al aprovechamiento de los datos, su recopilación, procesamiento, análisis e interpretación. La tecnología ha salvado innumerables vidas, mejorado los medicamentos y su tratamiento, creado dispositivos de apoyo para médicos y pacientes, actualizado procedimientos e incluso distribuyendo la organización administrativa para mejorar la calidad de vida de las personas.

Los servicios de salud vistos como un derecho humano que todo gobierno debe proteger bajo un conjunto de principios, normas e instituciones (también conocidos como sistemas de salud) se fueron constituyendo a partir de la creación de la Organización Mundial de la Salud por la Conferencia Sanitaria Internacional celebrada en Nueva York del 19 de junio al 22 de julio de 1946 (firmada el 22 de julio del mismo año por 61 países y entrando en vigor el 7 de abril de 1948)⁴⁶⁹.

Resulta sorprendente ver que fue hasta 1948 cuando los gobiernos del mundo reunidos en Nueva York, deciden entender a la salud como un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afectaciones o

⁴⁶⁹ Asamblea Mundial de la Salud, Constitución de la Organización Mundial de la Salud, Documentos básicos, suplemento de la 45ª edición, octubre de 2006, WHO, Ginebra, consúltase en: https://www.who.int/governance/eb/who_constitution_sp.pdf

enfermedades. Ante la grave crisis generada por la pandemia del COVID-19, creemos relevante recordar parte del contenido de la Constitución de la Organización Mundial de la Salud (principios y funciones relevantes):

“LOS ESTADOS parte en esta Constitución declaran, de conformidad con la Carta de las Naciones Unidas, **que los siguientes principios son básicos para la felicidad, las relaciones armoniosas y la seguridad de todos los pueblos: El goce del grado máximo de salud que se pueda lograr es uno de los derechos fundamentales de todo ser humano sin distinción de raza, religión, ideología política o condición económica o social.** La salud de todos los pueblos es una **condición fundamental para lograr la paz y la seguridad, y depende de la más amplia cooperación de las personas y de los Estados. Los resultados alcanzados por cada Estado en el fomento y protección de la salud son valiosos para todos. [...] El desarrollo saludable del niño es de importancia fundamental; la capacidad de vivir en armonía en un mundo que cambia constantemente es indispensable para este desarrollo. La extensión a todos los pueblos de los beneficios de los conocimientos médicos, psicológicos y afines es esencial para alcanzar el más alto grado de salud. [...] Los gobiernos tienen responsabilidad en la salud de sus pueblos, la cual sólo puede ser cumplida mediante la adopción de medidas sanitarias y sociales adecuadas. [...] Artículo 1. La finalidad de la Organización Mundial de la Salud... será **alcanzar para todos los pueblos el grado más alto posible de salud.** Artículo 2. Para alcanzar esta finalidad, las funciones de la Organización serán: a) actuar como **autoridad directiva y coordinadora en asuntos de sanidad internacional;** b) establecer y mantener colaboración eficaz con las Naciones Unidas, los organismos especializados, las administraciones oficiales de salubridad, las agrupaciones profesionales y demás organizaciones que se juzgue convenientes; **c) ayudar a los gobiernos, a su solicitud, a fortalecer sus servicios de salubridad; d) proporcionar ayuda técnica adecuada y, en casos de emergencia, prestara los gobiernos la cooperación necesaria que soliciten, o acepten;** [...] f) establecer y mantener los servicios administrativos y técnicos que sean necesarios, inclusive los *epidemiológicos* y de estadística; **g) estimular y adelantar labores destinadas a suprimir enfermedades epidémicas, endémicas y otras;** [...] k) **proponer convenciones, acuerdos y reglamentos y hacer recomendaciones referentes a asuntos de salubridad internacional, así como desempeñar las funciones que en ellos se asignen a la Organización y que estén de acuerdo con su finalidad;****

n) promover y realizar investigaciones en el campo de la salud; o) promover el mejoramiento de las normas de enseñanza y adiestramiento en las profesiones de salubridad, medicina y afines; p) estudiar y dar a conocer, con la cooperación de otros organismos especializados, cuando fuere necesario, técnicas administrativas y sociales que afecten la salud pública y la asistencia médica desde los puntos de vista preventivo y curativo, incluyendo servicios hospitalarios y el seguro social; t) establecer normas uniformes de diagnóstico, según sea necesario; u)

desarrollar, establecer y promover normas internacionales con respecto a productos alimenticios, biológicos, farmacéuticos y similares...”⁴⁷⁰

Antes de esta fecha, la labor del médico, la función que realizaban los hospitales y la responsabilidad de los gobiernos tenía una concepción distinta, de manera que ningún documento protector de derechos fundamentales del siglo XVIII o XIX pudo contener algo similar a la definición de 1948.

Las declaraciones de derechos y las constituciones de los Estados (siglos XVIII y XIX), solo contemplaban derechos civiles y políticos. El derecho a la salud, como parte de los derechos económicos, sociales y culturales (segunda generación Derechos Humanos), quedó plasmado en la Declaración Universal de los Derechos Humanos de 1948 y se desarrolla con los pactos internacionales⁴⁷¹ para constituir un entramado normativo que se interrelaciona en busca de su materialización.⁴⁷²

El artículo 25 de la ***Declaración Universal de los Derechos Humanos***, menciona a la salud como parte del derecho a un nivel de vida adecuado⁴⁷³ y es reconocido con más detalle en el *Pacto Internacional de Derechos Económicos, Sociales y Culturales*, aprobado el 16 de diciembre de 1966:

⁴⁷⁰ Asamblea Mundial de la Salud, Constitución de la Organización Mundial de la Salud, Documentos básicos, suplemento de la 45ª edición, octubre de 2006, WHO, Ginebra, consúltese en: https://www.who.int/governance/eb/who_constitution_sp.pdf

⁴⁷¹ La Declaración Universal de los Derechos Humanos con los pactos (de derechos civiles y políticos; económicos, sociales y culturales) y sus protocolos facultativos se conoce como “Carta Internacional de los Derechos Humanos”. Cfr. El derecho internacional de los Derechos Humanos, Oficina del Alto Comisionado de Derechos Humanos de las Naciones Unidas, consúltese en: <https://www.ohchr.org/sp/professionalinterest/pages/internationallaw.aspx#:~:text=La%20Declaraci%C3%B3n%20Universal%2C%20junto%20con,Carta%20Internacional%20de%20Derechos%20Humanos%E2%80%9D>.

⁴⁷² El derecho a la salud visto como la posibilidad que tiene la personas de disfrutar en nivel más alto posible de salud física, mental y social y no solo la ausencia de afectaciones o enfermedades va más allá de estar sano y no puede ser visto como un objetivo programático a largo plazo. A pesar de las limitaciones de recursos, se debe ver como un derecho de atención inmediata, en el que los estados deben de garantizar un nivel mínimo de atención y suministro de condiciones para la población. Ningún Estado puede justificar el incumplimiento por falta de recursos, incluso cuando éstos sean escasos. Cfr. *Observación general N.º 14 (2000) sobre el derecho a la salud, aprobada por el Comité de Derechos Económicos, Sociales y Culturales*. Consúltese en: <https://www.acnur.org/fileadmin/Documentos/BDL/2001/1451.pdf>

⁴⁷³ “Toda persona tiene derecho a un nivel de vida adecuado que le asegure, así como a su familia, la salud y en especial la alimentación, el vestido, la vivienda, la asistencia médica y los servicios sociales necesarios; tiene asimismo derecho a los seguros en caso de desempleo, enfermedad, invalidez, viudez, vejez u otros casos de pérdida de sus medios de subsistencia por circunstancias independientes de su voluntad.”

Artículo 12

1. Los Estados Parte en el presente Pacto reconocen el derecho de toda persona al disfrute del más alto nivel posible de salud física y mental.
2. Entre las medidas que deberán adoptar los Estados Parte en el Pacto a fin de asegurar la plena efectividad de este derecho, figurarán las necesarias para:
 - a) La reducción de la mortinatalidad y de la mortalidad infantil, y el sano desarrollo de los niños;
 - b) El mejoramiento en todos sus aspectos de la higiene del trabajo y del medio ambiente;
 - c) La prevención y el tratamiento de las enfermedades epidémicas, endémicas, profesionales y de otra índole, y la lucha contra ellas;
 - d) La creación de condiciones que aseguren a todos asistencia y servicios médicos en caso de enfermedad.⁴⁷⁴

Respecto al inciso c) el Comité de Derechos Económicos, Sociales y Culturales de la ONU en su observación general No. 14 (2000) señala que el derecho a tratamiento comprende “la creación de un sistema de atención médica urgente en los casos de accidentes, *epidemias* y peligros análogos para la salud, así como la prestación de socorro en casos de desastre y de ayuda humanitaria en situaciones de emergencia. La lucha contra las enfermedades tiene que ver con los esfuerzos individuales y colectivos de los Estados para facilitar, entre otras cosas, las tecnologías pertinentes, el empleo y la mejora de la vigilancia *epidemiológica* y la reunión de datos desglosados, la ejecución o ampliación de programas de vacunación y otras estrategias de lucha contra las enfermedades infecciosas.”⁴⁷⁵

En el caso de América Latina y la Unión Europea, encontramos legislación internacional que regula este tema y lo enfoca como una exigencia que se debe atender por los Estados para el mejoramiento de la salud de las personas:

⁴⁷⁴ Pacto Internacional de Derechos Económicos, Sociales y Culturales, Adoptado y abierto a la firma, ratificación y adhesión por la Asamblea General en su resolución 2200 A (XXI), de 16 de diciembre de 1966, consúltese en: <https://www.ohchr.org/sp/professionalinterest/pages/cescr.aspx>

⁴⁷⁵ Cfr. El derecho al disfrute del más alto nivel posible de salud: 11/08/2000.E/C.12/2000/4, CESCR OBSERVACION GENERAL 14. Consúltese en: <https://www.acnur.org/fileadmin/Documentos/BDL/2001/1451.pdf>

1. Organización Panamericana de la Salud (OMS).- La OMS considera que las desigualdades en la era digital deben ser reducidas y hace un llamado a la acción conjunta de las naciones bajo ocho principios que reconocen a la transformación digital como algo irreversible y benéfica en la que los estados latinoamericanos deben trabajar coordinadamente para motivar la cooperación supranacional del conocimiento e iniciar un proceso de alfabetización digital como una política pública desde las primera etapas de la educación, y así, fortalecer las bases fundacionales y los mecanismos de gobernanza para que las inversiones y acciones sean sostenibles, efectivas e informadas, cuidando la Ética, la inclusión social y los Derechos Humanos.⁴⁷⁶ Los ocho principios se incorporan en la siguiente tabla:

⁴⁷⁶ Organización Panamericana de la Salud, Ocho principios rectores de la transformación digital del sector de la salud. Un llamado a la acción panamericana OPS/EIH/IS/21-2004, Washington, 2021.

8 principios rectores de la transformación digital del sector de la salud

1	Conectividad universal 	Asegurar la conectividad universal en el sector de la salud para 2030
2	Bienes públicos digitales 	Cocrear bienes de salud pública digitales por un mundo más equitativo
3	Salud digital inclusiva 	Acelerar hacia una salud digital inclusiva con énfasis en los más vulnerables
4	Interoperabilidad 	Implementar sistemas de información y salud digital interoperables, abiertos y sostenibles
5	Derechos humanos 	Transversalizar los derechos humanos en todas las áreas de la transformación digital en salud
6	Inteligencia artificial 	Participar en la cooperación mundial sobre inteligencia artificial y cualquier tecnología emergente
7	Seguridad de la información 	Establecer mecanismos de confianza y seguridad de la información en el entorno digital de la salud pública
8	Arquitectura de salud pública 	Diseñar la arquitectura de la salud pública en la era de la interdependencia digital

Fuente: Organización Panamericana de la Salud, Ocho principios rectores de la transformación digital del sector de la salud. Un llamado a la acción panamericana OPS/EIH/IS/21-2004, Washington, 2021.

Actualmente, los retos que se tienen en salud para la próxima década requieren voluntad política, presupuestos, gente experta en los temas tecnológicos, participación del gobierno, las empresas y los científicos y van encaminados a que los países cuenten con lo mínimo indispensable para:

- Crear registros y llevar a cabo acreditaciones, clasificaciones, controles y monitoreo de prácticas, desarrollos, procesos y evaluación de resultados, con especial énfasis en la protección legal de las personas y entidades que intervengan (operadores, técnicos y profesionales, prestadores de servicios, intermediarios, comercializadores, etc.), sus productos, actividades, servicios y sistemas en su vinculación con la salud digital.
- Entablar alianzas y mecanismos sostenibles de intercambio de datos, información, conocimiento, inversión y financiamiento público-privado, que

incluyan fondos gubernamentales y bancos multilaterales en conjunción con el sector privado.

- c. Impulsar los marcos jurídicos y éticos de códigos abiertos, datos abiertos, modelos abiertos de inteligencia artificial, normas y contenidos abiertos. Especialmente, actualizar y reforzar los marcos de propiedad intelectual sobre el uso y acceso a los bienes públicos digitales, para facilitar su adquisición y aplicación en regiones de bajos recursos.
- d. Diseñar y adoptar nuevos mecanismos de cooperación entre el sector privado, el sector público, la sociedad civil y el mundo universitario.
- e. Incorporar los criterios de género, la perspectiva intercultural y los principios de equidad y solidaridad en las acciones relacionadas con la agenda de inclusión de la salud digital.
- f. Promover la inclusión de la salud digital centrada en las personas, asegurando que la población conozca sus derechos y responsabilidades en relación con la seguridad, la privacidad y la fiabilidad de los datos de salud.
- g. Incorporar las siguientes cuatro dimensiones en los planes nacionales y las políticas públicas de los sistemas de información y salud digital: 1) gestión y gobernanza de los sistemas de información; 2) gestión de datos y tecnologías de la información; 3) información y gestión del conocimiento e innovación; y 4) integración y convergencia digital, es decir, la posibilidad de acceder a un mismo contenido desde diferentes dispositivos.
- h. Consolidar una infraestructura de intercambio de datos abiertos e información crítica centrado en criterios éticos y de ciberseguridad en los flujos de información.
- i. Adoptar un programa de alfabetización digital basado en las necesidades detectadas y atendiendo a los diferentes contextos, con miras a reducir las inequidades.
- j. Formular instrumentos normativos apropiados, que ponderen el interés público frente a los datos individuales y masivos en salud, es decir, que busquen un equilibrio entre proteger la salud pública como bien social y los datos sensibles de los particulares, como ocurre con las tecnologías de notificación en vigilancia epidemiológica.
- k. Promover el desarrollo o modernización de ecosistemas legislativos sostenibles en salud digital, y en particular en relación con las leyes de acceso a información pública, asegurando la privacidad de los datos sensibles del ciudadano, con su consentimiento expreso.
- l. Asegurar el derecho de verificación, la validación de algoritmos y la evaluación de la tecnología digital en términos de seguridad, eficacia y sostenibilidad.
- m. Contar con instrumentos normativos que regulen el tratamiento y el acceso a los datos de salud desde los ejes de privacidad, confidencialidad y seguridad de la información.
- n. Establecer políticas públicas que incorporen un plan de seguridad y protección de los datos de salud, definiendo perfiles de acceso en función de las acciones que deba realizar el usuario.
- o. Habilitar servicios centralizados de certificación de seguridad de datos sensibles de salud mediante tecnologías de certificación de bloques (blockchain), etc.
- p. Actualizar las agendas digitales como una política pública que contenga el marco regulatorio necesario para las aplicaciones digitales en salud.
- q. Contemplar el establecimiento de tecnologías necesarias para permitir el intercambio de datos, por ejemplo, un bus de interoperabilidad, con servicios de consulta a bases de datos estandarizadas de beneficio y uso nacional.
- r. Definir procesos para evaluar las tecnologías emergentes vinculadas a los macrodatos (big data), el aprendizaje automático, la inteligencia artificial y las ciencias “ómicas”, entre otros.

- s. Formar equipos de gestión del cambio que acompañen la transformación digital, específicamente en el entorno de la salud, atendiendo a las necesidades de los distintos actores (pacientes, profesionales de la salud, personal administrativo, líderes y coordinadores, etc.).⁴⁷⁷

De gran importancia es el objetivo 6, en el que directamente se busca que los estados participen en la cooperación mundial sobre inteligencia artificial y cualquier tecnología emergente, como parte de la política de salud en el que la esfera individual de las personas y sus derechos se vincule con la dimensión social del derecho a la salud y, en conjunto, se vean como una condición humana. En este sentido, la Organización Panamericana de la Salud hace un llamado a todos los países para:

- a. Prevenir sesgos algorítmicos y garantizar la inclusión social desde el diseño de los sistemas de inteligencia artificial hasta su potencial aplicación en todos los niveles económicos, sociales y geopolíticos de los países de la Región.
- b. Promover fondos de inversiones multinacionales para que los países más vulnerables no se vean privados de los beneficios de las tecnologías emergentes y la inteligencia artificial en el campo de la salud pública.
- c. Asegurar el doble papel del Estado, como garante de los Derechos Humanos y promotor del goce de esos derechos, a la hora de desplegar la inteligencia artificial en el campo de la salud.
- d. Establecer mecanismos de coordinación formal que promuevan y aseguren el entendimiento cooperativo y constructivo entre segmentos públicos y privados, de gobernanza, universidad, industria y ciudadanía, a nivel regional y global.
- e. Construir un sistema de innovación en inteligencia artificial abierto y cooperativo, promoviendo el desarrollo de códigos de acceso abierto a algoritmos y aplicaciones.
- f. Acelerar los marcos ético-legales, la capacitación y encuentro de talento de inteligencia artificial de alto nivel.
- g. Construir una disciplina académica de inteligencia artificial mediante la formación de equipos multidisciplinarios (p. ej., profesionales de la salud, científicos de datos, profesionales legales) que fomenten temas de transparencia, explicabilidad y auditoría algorítmica, entre otros.⁴⁷⁸

En todo esto destaca una realidad, la necesidad de crear tecnología acorde a la idiosincrasia cultural de las naciones latinoamericanas en el que prevalezca la no discriminación, la inclusión con base en la equidad y el respeto a los Derechos Humanos y la capacidad de adaptabilidad al cambio en cada nación. Nuestros pueblos tienen un alto problema de analfabetismo digital que se convierte en un obstáculo serio para el avance de cualquier proyecto tecnológico -

⁴⁷⁷ *Idem.*

⁴⁷⁸ *Idem.*

independientemente del alcance o profundidad que se desee- y deriva de los altos índices de pobreza, baja calidad educativa, entornos sociales inequitativos y rodeados de corrupción, entre otros.

2. Unión Europea (UE). - En la UE existe el Convenio para la Protección de los Derechos Humanos y la Dignidad del Ser Humano con respecto a las aplicaciones de la Biología y la Medicina⁴⁷⁹ o Convenio de Oviedo, publicado el 4 de abril de 1997, en dicha ciudad española. En las áreas de la biología y la medicina la Inteligencia Artificial ha cobrado gran relevancia y si bien el tema no se menciona como tal, los avances más relevantes del genoma humano y la bioÉtica se están construyendo con *Machine Learning*⁴⁸⁰.

El Convenio tiene por objeto proteger al ser humano “en su dignidad y su identidad garantizar a todas las personas, sin discriminación alguna, el respeto a su integridad y a sus demás derechos y libertades fundamentales con respecto a las aplicaciones de la biología y la medicina.”⁴⁸¹

Una intervención en el ámbito de la sanidad dice el artículo 5, solo podrá efectuarse después de que la persona afectada haya dado su libre e inequívoco consentimiento. La persona que acepte debe recibir previamente la información adecuada acerca de la finalidad y la naturaleza de la intervención, así como sobre sus riesgos y consecuencias y, en cualquier momento, podrá retirar su consentimiento voluntario. El interés y el bienestar del ser humano deberán prevalecer sobre el interés exclusivo de la sociedad o la ciencia⁴⁸². El cuerpo humano y todas sus partes, por ningún motivo serán “fuente de aprovechamiento”.⁴⁸³

En materia de “privacidad” y “experimentación científica” los artículos 10 y 16 de la Convención de Oviedo contiene reglas básicas que los estados deben contemplar:

⁴⁷⁹ El Convenio de Oviedo entró en vigor el 1 de diciembre de 1999, consúltese en: <https://archivos.juridicas.unam.mx/www/bjv/libros/5/2290/37.pdf>

⁴⁸⁰ Poner ejemplos

⁴⁸¹ Artículo 1 del Convenio de Oviedo.

⁴⁸² Artículo 2 del Convenio de Oviedo.

⁴⁸³ Artículo 21 del Convenio de Oviedo.

Artículo 10. Vida privada y derecho a la información

1. Toda persona tendrá derecho a que se respete su vida privada cuando se trate de informaciones relativas a su salud.

2. Toda persona tendrá derecho a conocer toda información obtenida respecto a su salud. No obstante, deberá respetarse la voluntad de una persona de no ser informada.

[...]

Artículo 16. Protección de las personas que se presten a un experimento

No podrá hacerse ningún experimento con una persona, a menos que se den las siguientes condiciones:

i. Que no exista un método alternativo al experimento con seres humanos de eficacia comparable.

ii. Que los riesgos en que pueda incurrir la persona no sean desproporcionados con respecto a los beneficios potenciales del experimento.

iii. Que el proyecto de experimento haya sido aprobado por la autoridad competente después de haber efectuado un estudio independiente acerca de su pertinencia científica, comprendida una evaluación de la importancia del objetivo del experimento, así como un estudio multidisciplinar de su aceptabilidad en el plano ético.

iv. Que la persona que se preste a un experimento esté informada de sus derechos y las garantías que la ley prevé para su protección.

v. Que el consentimiento a que se refiere el artículo 5 se haya otorgado expresa y específicamente y esté consignado por escrito. Este consentimiento podrá ser libremente retirado en cualquier momento.

Las limitaciones en el tema científico parten del respeto al principio de la autonomía de la voluntad y la protección de los Derechos Humanos de las personas. Los experimentos con personas siempre tendrán un riesgo y un beneficio, esto se debe ponderar por las autoridades previo a la autorización del experimento. Además de la Convención de Oviedo, se aprobaron cuatro protocolos adicionales que sirven de guías operativas: 1. Por el que se prohíbe la clonación de seres humanos; 2. Sobre trasplante de órganos y tejidos de origen humano; 3. Relativo a la investigación biomédica; y, 4. Sobre pruebas genéticas para fines de salud.

En esta revisión del panorama normativo internacional destaca la labor del Comité de Ministros del Consejo de Europa y su interés respecto a la protección de los Derechos Humanos de las personas y las posibles afectaciones que generen autoridades al momento de tomar decisiones con base en inteligencia artificial o en

Machine Learning. El Consejo de Europa está consciente de que existen sistemas para la toma de decisiones asistidos por computadora o habilitados por Inteligencia Artificial que pueden incorporar un código, al que se confía la toma de decisiones, que puede reducir o negar erróneamente derechos y beneficios a las personas, lo que incidiría en una afectación a sus Derechos Humanos. Si bien la inteligencia artificial y el aprendizaje automático ofrecen grandes ventajas, se debe garantizar que las aplicaciones públicas tomen en cuenta los valores éticos y se apliquen a todos sin causar ninguna discriminación.

Para ello, el desarrollo no regulado de sistemas para la toma de decisiones automatizadas o asistidas por computadora, utilizadas por agencias de gobierno, sumado a la falta de transparencia y a su incorporación a la administración de los servicios sociales plantea riesgos naturales. Esto puede generar un impacto negativo para los miembros de la comunidad que se encuentran en situación de vulnerabilidad. En particular, preocupa al Consejo de Europa que pueden reproducirse patrones de discriminación arraigados en algunos sectores de la sociedad, en lo que respecta a las mujeres, niñas, niños o trabajadores con menores salarios o menos cualificados.

La IA basada en Machine Learning – Deep Learning y las decisiones que se tomen con su utilización, deben ser compatibles respecto de las disposiciones de la Carta Social Europea⁴⁸⁴. La Carta regula en sus artículos 11 y 12, la protección de la salud y el derecho a la seguridad social:

“Artículo 11. Derecho a la protección de la salud. Para garantizar el ejercicio efectivo del derecho a la protección de la salud, las partes contratantes se comprometen a adoptar, directamente o en cooperación con organizaciones públicas o privadas, medidas adecuadas para entre otros fines: 1. Eliminar, en lo posible, las causas de una salud deficiente. 2. Establecer servicios educacionales y de consulta dirigidos a la mejora de la salud y a estimular el sentido de responsabilidad individual en lo concerniente a la misma. 3. Prevenir, en lo posible, las enfermedades epidémicas, endémicas y otras.

Artículo 12. Derecho a la Seguridad Social. Para garantizar el ejercicio efectivo al derecho a la Seguridad Social, las partes contratantes se comprometen: 1. A establecer o mantener un régimen de Seguridad Social. 2. A mantener el régimen de Seguridad Social en un nivel

⁴⁸⁴ Cfr. Carta Social Europea de 1961, Turín, 18 de octubre de 1961. Consejo de Europa (Estrasburgo). [consúltase en: https://www.acnur.org/fileadmin/Documentos/BDL/2002/1934.pdf?file=t3/fileadmin/Documentos/BDL/2002/1934](https://www.acnur.org/fileadmin/Documentos/BDL/2002/1934.pdf?file=t3/fileadmin/Documentos/BDL/2002/1934)

satisfactorio, equivalente, por lo menos, al exigido para la ratificación del Convenio Internacional del Trabajo (número 102) sobre normas mínimas de Seguridad Social. 3. A esforzarse por elevar progresivamente el nivel del régimen de Seguridad Social. 4. A adoptar medidas, mediante la conclusión de los oportunos acuerdos bilaterales o multilaterales, o por otros medios, sin perjuicio de las condiciones establecidas en esos acuerdos, encaminadas a conseguir: a. La igualdad de trato entre los nacionales de cada una de las partes contratantes y los de las demás partes en lo relativo a los derechos de Seguridad Social, incluida la conservación de las ventajas concedidas por las leyes de Seguridad Social, sean cuales fueren los desplazamientos que las personas protegidas pudieren efectuar entre los territorios de las partes contratantes. b. La concesión, mantenimiento y restablecimiento de los derechos de Seguridad Social, por medios tales como la acumulación de los períodos de seguro o de empleo completados de conformidad con la legislación de cada una de las partes contratantes.”

En este sentido, no solo se puede violar el derecho a la salud y los derechos sociales sino todos los derechos que se encuentren interrelacionados o interconectados.

El uso de la IA basada en Machine Learning – Deep Learning puede ir en contra de la Recomendación Rec (2000) 3 del Comité de Ministros a los Estados miembros sobre el derecho a la satisfacción de las necesidades materiales básicas de las personas en situaciones de extrema dificultad, mediante la cual el Comité de Ministros convocó a todos que los Estados miembros a reconocer un derecho individual, universal y exigible a la satisfacción de las necesidades materiales básicas (como mínimo: alimentación, vestido, vivienda y atención médica básica) para las personas en situaciones de extrema carencia.

En la sesión del 17 de marzo de 2021, el Comité de Ministros de los Estados Miembros del Consejo de Europa emitieron una declaración recomendando a los Estados miembros atender lo siguiente:

- los posibles **riesgos** para los Derechos Humanos, incluidos los derechos sociales, que podrían derivarse del **uso en la toma de decisiones asistida por ordenador o mediante IA por parte de las autoridades en el ámbito de los servicios sociales**, cuando **no se garantiza una supervisión cualificada**, en particular en lo que respecta a la gestión, atribución o revocación de derechos, asistencia y beneficios relacionados;
- la **necesidad de garantizar** que los **sistemas de toma de decisiones** asistidos por ordenador o habilitados por IA **se desarrollen y apliquen de acuerdo con los principios de seguridad jurídica, legalidad, calidad de los datos, no discriminación y transparencia**. Los conocimientos y habilidades de los usuarios de estos sistemas deben mejorarse periódicamente;
- la **necesidad de supervisión humana** de las decisiones asistidas por computadora o habilitadas por IA **para mitigar y / o evitar errores en la gestión, atribución o revocación**

de derechos, asistencia y beneficios relacionados que podrían amplificar las desventajas y / o la privación de derechos;

- la necesidad de acuerdos efectivos para **proteger a las personas vulnerables de daños graves o irreparables, incluida la indigencia, la pobreza extrema o la falta de vivienda**, como resultado de la implementación de decisiones asistidas por computadora o habilitadas por IA en el área de los servicios sociales;

- la **necesidad de procesos efectivos de responsabilidad y rendición de cuentas** para los actores de la IA que diseñan, desarrollan, implementan o evalúan sistemas de IA cuando no se respetan las normas legales o se produce un daño injusto;

- un enfoque proactivo con miras a **garantizar que los afectados por decisiones asistidas** por computadora o habilitadas por IA en el ámbito de los servicios sociales, en particular las **personas** en una situación de **extrema** privación o **vulnerabilidad** puedan hacer valer efectivamente sus derechos y buscar reparación. En este sentido, **se entregará a los interesados una explicación adaptada al contexto y audiencia específicos**. Dicha explicación debe, al menos, proporcionar los elementos necesarios para **permitir que una persona comprenda y impugne una decisión que haya sido informada o tomada por un sistema de inteligencia artificial y que afecte su posición legal** o su vida de manera sustantiva.⁴⁸⁵

Lo relevante del comunicado está en la obligación que se está generando en todos los gobiernos de la Unión Europea, que utilicen sistemas de inteligencia artificial o de aprendizaje automático de prevenir riesgos que afecten, por su uso, a las personas en sus Derechos Humanos. Visto esto desde un enfoque de contratación de proveedores o del acto administrativo de licitación o adjudicación

⁴⁸⁵ Por tanto, el Comité de Ministros de los Estados miembros pide que se atiendan:

-the possible risks to human rights, including social rights, that might follow from the use of computer-assisted or AI-enabled decision making by public authorities in the area of social services, when no qualified supervision is guaranteed, in particular as regards the management, attribution or revocation of entitlements, assistance and related benefits;-the need to ensure that computer-assisted or AI-enabled decision-making systems are developed and implemented in accordance with the principles of legal certainty, legality, data quality, non-discrimination and transparency. The knowledge and skills of the users of these systems are to be enhanced regularly.

-the need for human oversight of computer-assisted or AI-enabled decisions in order to mitigate and/or avoid errors in the management, attribution or revocation of entitlements, assistance and related benefits which could amplify disadvantages and/or disenfranchisement.

- the need for effective arrangements to protect vulnerable persons from serious or irreparable harm, including destitution, extreme poverty, or homelessness, as a result of the implementation of computer-assisted or AI-enabled decisions in the area of social services.

- the need for effective responsibility and accountability processes for AI actors designing, developing, deploying or evaluating AI systems when legal norms are not respected, or any unjust harm occurs.

- a proactive approach with a view to ensuring that those affected by computer-assisted or AI-enabled decisions in the area of social services, in particular persons in a situation of extreme deprivation or vulnerability, can effectively assert their rights and seek remedies. In this regard, an explanation tailored to the specific context and audience shall be delivered to those concerned. Such explanation should at least provide the necessary elements to allow an individual to understand and challenge a decision that has been informed or made by an AI system, and that affects his or her legal position, or his or her life in a substantive manner.

directa, se debe incluir -adicional a los requisitos actuales para la adquisición de sistemas de este tipo- una evidencia de que el sistema o programa no afecta los Derechos Humanos de las personas. La obligación de las autoridades, en el ámbito de los servicios sociales, se vincula con el resto de los Derechos Humanos con los que se interrelacionen. Y esto debe ser un requisito previo a la contratación o en el cuerpo del instrumento contractual que se utilice para formalizar la prestación del servicio, cuidando los principios de seguridad jurídica, legalidad, calidad de los datos, no discriminación y transparencia.

Otro punto importante será al momento de su utilización, es decir, ya iniciado el contrato de prestación del servicio. Aquí habrá la obligación del propio gobierno de supervisar el sistema y no dejarlo actuar sin la intervención del ser humano para mitigar y / o evitar errores en la gestión, atribución o revocación de derechos, asistencia y beneficios relacionados que podrían amplificar las desventajas y / o la privación de derechos.

Y finalmente, la creación de un procedimiento legal efectivo -ante autoridad judicial- de responsabilidad algorítmica y una metodología de evaluación de riesgos que sea sometida a una auditoría algorítmica. El primero, para aquellas personas que se sientan afectadas en sus derechos. El segundo, para que el gobierno y la empresa, puedan mitigar riesgos, corregir y mejorar su uso, evitándose así -en la medida de lo posible- violaciones de Derechos Humanos.

En la siguiente sección, analizaremos un quinto caso relacionado con el uso de la tecnología cognitiva en el que participa el ser humano, se toman decisiones por una máquina y hay procesamiento de datos bajo un análisis tridimensional. Nos preguntaremos 1) si las categorías actuales de la responsabilidad y la dogmática sirven para solucionar la problemática del uso de sistemas inteligentes que ayudan a la toma de decisiones. En caso de no ser suficientes para dar una solución, entonces nos preguntamos qué debemos añadir 2); y, finalmente, 3) si a la par de cualquiera respuesta anterior, debería haber una respuesta basada en la Ética.

Capítulo IV. Categorías de la responsabilidad jurídica, su aplicación a casos de sistemas de Inteligencia Artificial para la toma de decisiones y el uso de la Ética empresarial. Estudio del caso *Watson Health* de IBM

IV. Categorías de la responsabilidad, su aplicación a casos de sistemas de Inteligencia Artificial para la toma de decisiones y el uso de la Ética empresarial. Estudio del caso *Watson Health* de IBM

Esta sección tiene como objetivo particular aplicar los conocimientos adquiridos a lo largo de la investigación y utilizarlos para fijar la discusión en las categorías de la responsabilidad, su aplicación a casos de sistemas de Inteligencia Artificial para la toma de decisiones y el uso de la Ética empresarial con el fin de dar a las empresas un conjunto de indicadores de medición de riesgos generados por el uso de la IA. Buscamos con ello, dar a las empresas soluciones preventivas que puedan incorporarse a los programas de integridad corporativa o a los programas de cumplimiento legal para mitigar riesgos.

Esto se hará desde tres distintas dimensiones con el fin de ubicar los riesgos legales, tecnológicos y éticos que están presentes en cualquier empresa que utilice IA, pero que muchas veces no se mitigan por un desconocimiento de estos temas. Las dos primeras dimensiones, son ontológicamente idénticas, pero epistémicamente distintas; y la tercera, tiene que ver con las aportaciones de la Ética empresarial.

Las dos primeras dimensiones tienen la misma ontología ya que surgen de un mismo objeto de estudio que es la realidad denominada “Inteligencia Artificial” (representando personas, objetos y conceptos de un área de interés y sus relaciones) y la búsqueda de su regulación basada en valores, normas y hechos que están ocurriendo en la sociedad, pero epistémicamente son distintas ya que la primera la encontramos en el sistema jurídico (norma jurídica) y en la doctrina que le da soporte; y la segunda, la ubicamos en el ámbito de la Ingeniería y la ciencia de datos (norma no vigente, pero sí aplicable en ocasiones por las organizaciones que las han creado). Ambas, se relacionan con la medición de riesgos legales y tecnológicos y su mitigación. Todo riesgo tecnológico a su vez puede ser un riesgo legal y ético, pero no a la inversa.

En cambio, la tercera dimensión se relaciona con las virtudes que debe tener el empresario, el lograr un ambiente laboral bueno, el buscar cumplir principios, valores y reglas de comportamiento (integridad empresarial) y sus fines van más

allá de las dos dimensiones previas tanto en el tiempo como en la amplitud de los beneficios.

En la primera dimensión, encontramos la norma vigente y la dogmática que le da vida y que le permite enjuiciar por un hecho ilícito o por la comisión de un delito a cualquier persona que esté en los supuestos contenidos en la propia ley. Aquí nos preguntamos si las categorías vigentes de la responsabilidad jurídica y su dogmática sirven para solucionar los problemas que pueden surgir por el uso de sistemas como *Watson Health* y analizaremos qué salidas ofrece la legislación.

En la segunda dimensión, tiene relación con conceptos que aún no se encuentran vigentes ni el Derecho mexicano ni en muchos otros sistemas jurídicos. Esta dimensión se construyó en los laboratorios y actividades académicas de la Ingeniería, la ciencia de datos y la Ética de la IA¹, a través de propuestas de diversas organizaciones (IEEE, ACM, OCDE, UNESCO, entre otras), expertos independientes y académicos (como el de interpretabilidad y explicabilidad), y es la base para crear una nueva categoría de responsabilidad por el uso de sistemas autónomos de IA. Aquí nos preguntamos si es necesario ampliar las categorías de responsabilidad existentes con la que denominamos: responsabilidad algorítmica.

La responsabilidad algorítmica surge conjuntamente con otro concepto denominado Agente Artificial Moral Autónomo o Persona Electrónica o Persona Algorítmica. El uso de estos conceptos, a su vez permitirá dividir la responsabilidad que le corresponde a la máquina de la responsabilidad que les corresponde a los seres humanos, como la de: 1) los que producen los datos, 2) los que procesan los datos, 3) los que desarrollan los sistemas; 4) los que utilizan el sistema; 5) los que adquieren el sistema; 6) los que certifican un sistema; y 7) los que auditan el sistema. En un futuro, la búsqueda de la responsabilidad del ser humano y la

¹ Un ejemplo de nuevos conceptos es la llamada “Ética algorítmica” que se pregunta cómo la IA, aprendizaje máquina y robots interpretan los datos y de una “Ética de la IA”, que prioriza consideraciones para la creación de sistemas de IA, promoviendo estándares o difundiendo políticas globales, regionales o nacionales para su uso; también hay una “Ética de las máquinas”, que busca construir sistemas inteligentes que tomen decisiones autónomas y se pregunta cómo codificar la moral en las máquinas; y, hay una “IA moral”, dedicada a conseguir la cooperación máquina-ser humano para resolver problemas complejos con el uso de la moral; la IA es vista como un mecanismo para la mejora de los seres humanos. Cfr. Aníbal Monasterio Astobiza, “Ética algorítmica: Implicaciones Éticas de una sociedad cada vez más gobernada por algoritmos”, *Dilemata*, número 24, mayo 2017, pp. 185-217.

búsqueda de la responsabilidad de las máquinas será parte de la valoración que hagan jueces, peritos, fiscales, árbitros, etcétera.

Y, finalmente, la tercera dimensión tiene que ver con la Ética empresarial y busca regular conductas que pueden generar afectaciones a las personas (dentro y fuera de la organización). Aquí nos preguntamos si a la par de una respuesta jurídica debería haber una respuesta Ética; nosotros creemos que sí y que las actuales aportaciones de la Ética en la IA así como de la Ética empresarial proporciona el soporte adecuado para ello.

Llano lleva la Filosofía al mundo empresarial y construye lo que podemos llamar la Teoría de una Moral de Responsabilidad Empresarial con sus cuatro campos de actuación: 1) la responsabilidad del consecuente (o responsabilidad de las consecuencias de la acción); 2) la responsabilidad del antecedente (o responsabilidad de los principios); 3) la responsabilidad congruente (o responsabilidad de los proyectos anteriores de mi vida); y 4) la responsabilidad del trascendente (o de la misión a la que, con mi vida, he sido destinado). La importancia de utilizar esta teoría como ingrediente adicional a los controles de autoregulación o de integridad corporativa que las empresas tienen y como un complemento a sus programas de *Compliance*, les permitirá demostrar a la autoridad que la empresa ha hecho todo lo que está en sus manos para evitar una afectación derivada de sistemas de IA y evitará así, una posible sanción jurídica o reputacional.

Llano conocía perfectamente las distintas posturas de los abogados en el tema de la responsabilidad (Kant, Rawls, Dworkin, entre otros) y con su pensamiento las analizó y decidió construir categorías de responsabilidades específicas para la empresa. La empresa para Llano debe ser entendida no como un lugar de comercio de productos o servicios, sino como una comunidad de seres humanos con dignidad, inteligencia, fines en común y responsabilidades. No en vano se le conoce a Carlos Llano Cifuentes como el filósofo-empresario y el empresario-filósofo.

Las tres dimensiones, en conjunto, constituyen la metodológica para la medición de riesgos legales, tecnológicos y éticos que utiliza un catálogo de indicadores básicos como instrumento para que las empresas y los gobiernos

mitiguen sus riesgos, incorporando estos indicadores a su programa de integridad corporativa y a su programa de *Compliance* así como a los sistemas automatizados que utilizan para medir los riesgos.

Para demostrar la utilidad del Catálogo de indicadores básicos para la medición de riesgos derivados del uso de IA, haremos la problematización de un caso real en el que fallece una persona por lo que se conoce como Síndrome de compresión medular por cáncer de próstata metastásico². Esto proporciona elementos para comprender: a) las características y el uso actual de las categorías de responsabilidad vigentes (jurídicas, sin tecnología y con tecnología); b) las características y el uso de la nueva categoría de responsabilidad algorítmica (no vigente en varios sistemas jurídicos); y, c) la importancia del uso de la Ética empresarial.

Por tanto, la primera sección de este apartado se denomina *1) Las tres dimensiones de la responsabilidad*, y a manera de prolegómeno, un esbozo sobre el concepto responsabilidad y su evolución en el mundo occidental y, posteriormente, tres subsecciones: a) Categorías de la responsabilidad jurídica; b) Categorías de la Responsabilidad algorítmica y otros conceptos; c) Categorías de la responsabilidad moral. La segunda sección de este apartado se denomina: *2) El caso Watson Health y la aplicación de las tres dimensiones de la responsabilidad*, y consta también de dos subsecciones: a) Metodología para la medición de riesgos legales, éticos y tecnológicos; y, b) Problematicación de caso *Watson Health* de IBM.

1) Las tres dimensiones de la responsabilidad

Antes de entrar al primer apartado de esta sección, creemos necesario hacer un esbozo sobre el concepto de responsabilidad y su evolución en el mundo occidental seleccionando a tres filósofos que han estudiado el tema: Levy-Bruhl, Hart y Dworkin. Explicar de manera monográfica a estos autores, permite a quienes son

² American Cancer Society, ¿Cómo entender el cáncer avanzado, el cáncer metastásico y la metástasis en los huesos?, *Cáncer avanzado, cáncer metastásico y metástasis en los huesos*, 15 de diciembre de 2016, s/l, consúltase en: <https://www.cancer.org/es/tratamiento/como-comprender-su-diagnostico/cancer-avanzado/que-es.html>

ajenos al tema de la responsabilidad, vincularse con él y, por tanto, entender cualquier actualización o modificación normativa que se busque realizar en el sector laboral al que pertenezcan.

Lucien Levy-Bruhl. El filósofo y sociólogo francés Lucien Levy-Bruhl (1857-1939), en su obra *L'Idée de Responsabilité*, (1884) analiza el tema siguiendo las aportaciones de Descartes, Spinoza, Leibniz, Hume y Kant. Levy-Bruhl llega a la conclusión que existen dos formas de entender a la responsabilidad. Una primera, basada en un enfoque jurídico-objetivo; otra, tomada desde el enfoque de la Moral (y de la conciencia humana)³ y lo explica de la siguiente manera:

A medida que la noción de responsabilidad objetiva se difunde y se afianza en la mente de las personas, también debe destacarse el carácter moral de la persona humana: más que nunca debemos ceñirnos al principio de Kant, que expresa tan bien el respeto al que tiene derecho: "*Trata siempre a la humanidad, en ti y en la de los demás, como un fin y nunca como un medio*". El hombre, por así decirlo, es sólo una unidad social, pero la relación de esta unidad con el todo no es una simple relación numérica. Lleva en sí mismo un valor absoluto y, por tanto, inconmensurable. Esto, naturalmente, limita el derecho de la sociedad al individuo. Por ejemplo, la ley penal no puede castigarlo con miras a intimidar a otros, aunque tales castigos serían de evidente interés social. Sería hacer del hombre un instrumento, tratarlo como un medio y no como un fin. Debe rechazar esta intimidación, o producirla indiscriminadamente, como consecuencia de sanciones que, además, son legítimas.⁴

³ « La notion de responsabilité s'est dédoublée. D'une part la notion de responsabilité légale s'est dégagée de tout élément subjectif, pour devenir exclusivement objective. De l'autre, la notion de responsabilité morale, vide en elle-même, emprunte au devoir un contenu positif, et peut être représentée symboliquement dans la conscience. » [La noción de responsabilidad se ha dividido en dos. Por un lado, la noción de responsabilidad jurídica se ha liberado de todo elemento subjetivo, para tornarse exclusivamente objetiva. Por otro lado, la noción de responsabilidad moral, vacía en sí misma, toma prestada un contenido positivo del deber y puede representarse simbólicamente en la conciencia.] Cfr. L. Levy-Bruhl, *L'Idée de Responsabilité*, Paris, Librairie Hachette et Cie, 1884.

⁴ A mesure que la notion d'une responsabilité purement objective se répand et se fixe dans les esprits, le caractère moral de la personne humaine doit aussi être mis en relief : plus que jamais il faut s'attacher au principe de Kant, qui exprime si bien le respect auquel elle a droit. « *Traite toujours l'humanité, dans ta personne et dans celle d'autrui, comme une fin, et jamais comme un moyen.* » L'homme, si l'on veut, n'est qu'une unité sociale, mais le rapport de cette unité au tout n'est pas un simple rapport numérique. Elle porte en elle-même une valeur absolue, et par suite inconmensurable. Par-là se trouve naturellement limité le droit de la société sur l'individu. La loi pénale ne saurait par exemple le punir en vue d'intimider les autres, quand même de telles punitions seraient d'un intérêt social très évident. Ce serait faire de l'homme un instrument, le traiter comme un moyen, et non comme une fin. Il faut, ou qu'elle renonce à cette intimidation, ou qu'elle la produise indirectement, comme conséquence de peines d'ailleurs légitimes. Cfr. L. Levy-Bruhl, *L'Idée de Responsabilité*, Paris, Librairie Hachette et Cie, 1884. [El énfasis es nuestro]

Estos ingredientes están muy relacionados con la forma en cómo entendemos al ser humano o persona. Esa sustancia individual de naturaleza racional de la que hablaba Boecio⁵ y la manera en la que nos interrelacionamos como agentes morales, autónomos⁶, capaces de decidir lo bueno y lo malo y, en consecuencia, de definir nuestro propio destino.⁷ Los actos “voluntarios” de un hombre, por sí mismos son imputables a él y esa imputabilidad corresponde a la responsabilidad del agente que los realiza.

La noción de responsabilidad nos la da nuestra conciencia moral⁸ y la conciencia moral es tomada en cuenta por el Derecho para medir las consecuencias de los actos de las personas ya que las consecuencias que se encuentra en ley impactan a la persona. Levy-Bruhl complementa diciendo:

No hay idea en nuestras mentes más clara, en la superficie, que la idea de responsabilidad. Parece que nos lo da inmediatamente la conciencia, *con la idea del libre albedrío*: sabemos que somos libres, por intuición directa. Al menos este testimonio de conciencia es tan universal, tan espontáneo, que los partidarios del libre albedrío encuentran en él una prueba de su realidad objetiva. *Realidad o ilusión, el testimonio de la conciencia, como hecho, es indiscutible*. El hombre se cree un ser razonable y libre. *Siendo razonable*, es capaz de imaginar varias acciones como posibles en un momento dado, y también es capaz de juzgar el valor relativo de estas acciones: siendo libre, puede elegir la que le gusta realizar. Con esta facultad de actuar con pleno conocimiento de los hechos y de elegir libremente la conducta que quiere seguir, el hombre *se atribuye una verdadera causalidad respecto a sus actos*. Los actos

⁵ Para conocer el significado de los elementos de esta definición puede consultarse a Gabriel Martí Andrés, “Sustancia individual de naturaleza racional: el principio personificador y la índole del alma separada” en *Metafísica y Persona. Filosofía, conocimiento y vida*, Año 1 enero-junio 2009 Número 1, Universidad de Málaga, pp. 113-119 en: <https://revistas.uma.es/index.php/myp/article/view/2849> De Boecio: “Liber de persona et duabus naturis”: ML, LXIV, 1343: “Persona est rationalis naturae individua substantia”. También de Tomás de Aquino: *Summa theologiae* (S. Th.) I, q. 29, a. 1. Tomás de Aquino: S. Th. II-II, q. 23, a. 3, ad 3: “Ad tertium dicendum quod omne accidens secundum suum esse est inferius substantia, quia substantia est ens per se, accidens autem in alio”. La editoria Porrúa tiene una edición de *La consolación de la Filosofía*, (2004).

⁶ Autonomía, del griego autós (uno mismo) y nómos (ley): independencia, autogobierno; ser, para sí mismo, la propia ley. Cfr. José Ramón Ayllón, *Antropología filosófica*, Barcelona, Ariel, 2014, p. 261.

⁷ Jesús González Amuchastegui, *Autonomía, Dignidad y Ciudadanía. Una teoría de los Derechos Humanos*, Tirant lo Blanch, Valencia, 2004, p. 121 y ss. Kant dice que el hombre “al ser racional y, por tanto, perteneciente al mundo inteligible, no puede pensar nunca la causalidad de su propia voluntad sino bajo la idea de la libertad, pues la independencia de las causas determinantes del mundo sensible (independencia que la razón tiene siempre que atribuirse) es libertad. Con la idea de la *libertad* hallase, empero, inseparablemente unido el concepto de *autonomía*, y con éste el principio universal de la *moralidad*, que sirve de fundamento a la idea de todas las acciones de seres racionales, del mismo modo que la ley natural sirve de fundamento a todos los fenómenos. Emmanuel Kant, *Fundamentación de la Metafísica de las Costumbres*, Trad. Manuel García Morente, Puerto Rico, Edición de Pedro M. Rosario Barbosa, 2007, p. 65

⁸ Conciencia moral, “es la misma razón humana en tanto que juzga sobre la moralidad de los actos, sobre el bien y el mal. Manifiesta al sujeto la exigencia de una norma Ética objetiva: la ley natural.”. Cfr. José Ramón Ayllón, *Antropología filosófica*, Barcelona, Ariel, 2014, p. 263.

*voluntarios de un hombre le son imputables. La rendición de cuentas de los actos corresponde a la responsabilidad del que actúa. Por tanto, la noción de responsabilidad sugiere inmediatamente la conciencia.*⁹

La conciencia, el razonamiento y el libre albedrío, han sido utilizados para analizar la responsabilidad por siglos, tanto por filósofos y juristas como por neurocientíficos. Estos últimos están construyendo aportaciones a múltiples campos del quehacer humano (no solo en medicina o derecho)¹⁰.

Si tomamos el caso de los filósofos, a lo largo de la historia han emitido opiniones sobre la responsabilidad y su relación con la libertad, la voluntad, el libre albedrío y la conciencia, generándose dos posturas. Una primera, a favor de la creencia del libre albedrío y todos los elementos que derivan de la esencia del ser humano; y una segunda, en contra, considerándolo un producto de la imaginación que es inexistente (una ilusión generada por nuestro desconocimiento de las causas de nuestros deseos y voliciones). “El determinismo que rige el acontecer universal, y por ello también los procesos que tienen lugar en nuestros cerebros, que determinan a su vez todos nuestros procesos mentales, incluyendo la deliberación y la toma de decisiones, desvelan la idea de libre albedrío como una construcción ficticia de nuestra mente”¹¹.

⁹ Il n'y a pas d'idée dans notre esprit plus claire, en apparence, que l'idée de la responsabilité. Il semble qu'elle nous soit donnée immédiatement par la conscience, avec l'idée du libre arbitre : nous savons que nous sommes libres, par une intuition directe. Au moins ce témoignage de la conscience est-il si universel, si spontané, que les partisans du libre arbitre y trouvent une preuve de sa réalité objective. Réalité ou illusion, le témoignage de la conscience, comme fait, est incontestable. L'homme se croit un être raisonnable et libre. Étant raisonnable, il est capable de se représenter plusieurs actions comme possibles à un moment donné, et capable aussi de juger de la valeur relative de ces actions : étant libre, il peut choisir celle qu'il lui plaît de réaliser. Avec cette faculté d'agir en connaissance de cause et de choisir librement la conduite qu'il veut suivre, l'homme s'attribue une causalité véritable à l'égard de ses actes. Les actes volontaires d'un homme lui sont imputables et l'imputabilité des actes correspond à la responsabilité de celui qui agit. La notion de responsabilité nous est donc immédiatement suggérée par la conscience. Cfr. L. Levy-Bruhl, *L'idée de Responsabilité*, Paris, Librairie Hachette et Cie, 1884. [El énfasis en propio]

¹⁰ El profesor español Joaquín M. Fuster ha hecho aportaciones muy orientadoras sobre estos temas. Consúltese: *Cerebro y Libertad. Los cimientos cerebrales de nuestra capacidad para elegir*, México, Ediciones Culturales Paidós, 2019.

¹¹ Carlos J. Moya, “Prologo” en el libro escrito por Alfred R. Mele, *Libres, por qué a ciencia no ha rebatido la existencia del libre albedrío*, Valencia, Avarigani, 2017. Carlos J. Moya es experto en el estudio del libre albedrío, responsabilidad moral y Filosofía de la mente y de la acción. Puede consultarse también: *Libre albedrío. Un estudio filosófico* de editorial Cátedra (2017).

Desde los racionalistas, como Spinoza, Descartes, Copérnico, Galileo hasta Stephen Hawking y los neurocientíficos como Benjamín Libet¹², afirman que nuestras intenciones y decisiones conscientes no son las causas reales de nuestras acciones, sino, como éstas, meros efectos (epifenómenos) de sucesos neurológicos inconscientes.”¹³.

Entre los filósofos que consideran que el libre albedrío sí existe y que se debe tomar en cuenta para valorar la responsabilidad de una persona en determinado hecho, encontramos a Aristóteles para quien la virtud es fruto de la voluntad (haciendo referencia a la autodeterminación de las personas)¹⁴: “depende de nosotros realizar buenas y malas acciones, e igualmente el no realizarlas (y esto era el ser buenos o malos) ... dependerá de nosotros el ser virtuosos o viciosos...”.¹⁵ Ya que el hombre es un ser dotado de razón, cuenta con las condiciones de definir

¹² Para más detalles consúltese el libro de Benjamin Libet, *Mind Time: The Temporal Factor in Consciousness*, Harvard University Press, 2005. El trabajo de Libet se ha centrado en las relaciones temporales entre los eventos neuronales y la experiencia. Es famoso en parte por descubrir que inconscientemente decidimos actuar bien antes de pensar que hemos tomado la decisión de actuar. Este hallazgo tiene importantes implicaciones para uno de los problemas más profundos de la Filosofía y la psicología, a saber, el problema del libre albedrío. En un experimento, Libet pidió a las personas que movieran la muñeca en el momento que eligieran. Los participantes informaron tener la intención unos 200 milisegundos antes de que realmente comenzaran a moverse. Libet también midió el potencial de preparación, que se revela mediante la actividad registrada en el área motora del cerebro. Este potencial de preparación se produjo unos 550 milisegundos antes de que comenzara la acción. Los eventos cerebrales que produjeron el movimiento ocurrieron así antes de que el participante fuera consciente de haber tomado una decisión. Libet muestra que esta disparidad no se debe simplemente al tiempo adicional requerido para anotar e informar el tiempo. Señala que incluso si un movimiento fue iniciado por fuerzas inconscientes, hay tiempo suficiente para vetar un acto, una vez que uno es consciente de sus intenciones. Esta observación mantiene la puerta abierta a las nociones tradicionales de libre albedrío.

¹³ Libet y otros neurocientíficos niegan la existencia del libre albedrío al realizar ciertos experimentos para demostrar que las causas de nuestras acciones son meros sucesos físicos de nuestro cerebro de los que no somos conscientes. En el libro de R. Melé se concluye que aún nadie ha podido demostrar si existe o no existe el libre albedrío, sigue un espacio de duda y falta de consensos. Cfr. Alfred R. Mele, *Libres, por qué a ciencia no ha rebatido la existencia del libre albedrío*, Avarigani, Valencia, 2017. Benjamín Libet es considerado uno de los principales neurocientíficos que ha tratado de influir en la inexistencia del libre albedrío. Consúltese: Libre albedrío en: <http://www.sefaweb.es/libre-albedrio/>

¹⁴ En griego persona o “prosopon” era la máscara que usaban los actores en el teatro. En latín “persona”, servía para dar resonancia a la voz y para identificar el personaje que representaba (el que representa un papel y también la interioridad que actúa en nosotros tras la fachada corporal). Cfr. José Ramón Ayllón, *Antropología filosófica*, Ariel, Barcelona, 2014, p. 288.

¹⁵ Aristóteles, *Moral a Nicómaco*, Madrid, Alianza Editorial, 2005, Libro III, 5, 1113b.

su propia conducta. Sin su racionalidad sería guiado por sus pasiones, dominado por ellas. La razón es la que lo ayuda a reflexionar sobre sus pasiones.¹⁶

Otro filósofo es San Agustín, quien también se declara partidario del libre albedrío manifestando que “si el mal no es obra de la voluntad, absolutamente nadie debe ser reprendido o amonestado, y con la supresión de todo esto recibe un golpe mortal la ley cristiana y toda disciplina religiosa. Luego a la voluntad debe atribuirse la comisión del pecado. Y como no hay duda sobre la existencia del pecado, tampoco la habrá de esto, conviene saber: que el alma está dotada del libre albedrío de la voluntad. Pues juzgó Dios que así serían mejores sus servidores, si libremente le servían, cosa imposible de lograrse mediante un servicio forzado y no libre”.¹⁷

Emmanuel Kant (1724-1804) entiende que “los seres racionales llámense *personas* porque su naturaleza los distingue ya como fines en sí mismos, esto es, como algo que no puede ser usado meramente como medio, y, por tanto, limita en ese sentido todo capricho (y es un objeto del respeto). Éstos no son, pues, meros fines subjetivos, cuya existencia, como efecto de nuestra acción, tiene un valor para nosotros, sino que son fines objetivos, esto es, cosas cuya existencia es en sí misma un fin, y un fin tal, que en su lugar no puede ponerse ningún otro fin para el cual debieran ellas servir de medios, porque sin esto no hubiera posibilidad de hallar en parte alguna nada con valor absoluto; más, si todo valor fuera condicionado y, por tanto, contingente, no podría encontrarse para la razón ningún principio práctico supremo.”¹⁸

En este orden de ideas, podemos afirmar que la responsabilidad tiene relación con la capacidad y obligación de asumir las consecuencias de nuestros actos libres. Este ingrediente es inseparable de la libertad y la posibilidad de las personas de

¹⁶ En este tema de la libertad y el libre albedrío también consúltese a Boecio en su Libro quinto de *La consolación de la Filosofía* y Heleno Saña, *Breve tratado de Ética. Una introducción a la teoría de la moral*, Madrid, Almuzara, 2009.

¹⁷ San Agustín, “El pecado procede del libre albedrío”, *De la verdadera religión*, XIV. 27, en: https://www.augustinus.it/spagnolo/vera_religione/index2.htm

¹⁸ Emmanuel Kant, *Fundamentación de la Metafísica de las Costumbres*, Trad. Manuel García Morente, Puerto Rico, Edición de Pedro M. rosario Barbosa, 2007, p. 42 en: https://pmrb.net/books/kantfund/fund_metaf_costumbres_vD.pdf José Ramón Ayllón añadiría que la “dignidad de su persona se ha hecho derivar de su condición de inteligente, libre y responsable, y en último caso, de su condición de criatura divina”. Cfr. José Ramón Ayllón, *Antropología filosófica*, Barcelona, Ariel, 2014, p. 288.

elegir entre lo bueno y lo malo. La responsabilidad jurídica, no está separada de la responsabilidad moral, por el contrario, el concepto de responsabilidad moral le da vida al concepto de responsabilidad jurídica contenida en los códigos penales, civiles, administrativos. La responsabilidad jurídica contiene ingredientes de la responsabilidad moral. En particular: el libre albedrío, la capacidad de las personas, el entendimiento del acto u omisión, la conciencia del mismo; y a contrario sensu, la falta de responsabilidad deriva de la ausencia de una acción u omisión (voluntaria y consciente). Esto se va a comprobar cuando entremos al estudio del caso de análisis y veamos cómo la visión de responsabilidad moral de Carlos Llano se puede utilizar a la par de nuestras categorías actuales de responsabilidad y a la par del nuevo concepto de responsabilidad que se cree para regular la IA basada en Machine Learning – Deep Learning. En cambio, las categorías de responsabilidad jurídica vigentes quedan más acotadas o, por lo menos, con poca claridad para su aplicación a los casos de responsabilidad que surge por el uso de la IA basada en Machine Learning – Deep Learning, confirmándose que las bases de la responsabilidad jurídica evolucionan tanto como evolucione la responsabilidad moral y sus nuevos conceptos.

H. L. A. Hart. Continuando con la evolución histórica de las categorías de responsabilidad, encontramos a Herbert Lionel Adolphus Hart (1907-1992) con su obra *Punishment and Responsibility*¹⁹, en la cual el autor hace una compilación de artículos que redacta a lo largo de su carrera académica; una producción amplia y bastante apasionante sobre Derecho Penal, pena, castigo y responsabilidad que toma las aportaciones de Beccaria, Bentham, Kant, Locke, Mill y Austin.²⁰ Hart

¹⁹ *Punishment and Responsibility*, Oxford, Clarendon Press, 1968. En 2019, la editorial Marcial Pons publica la edición traducida por Jacobo Barja de Quiroga y León García-Comendador Alonso: Castigo y responsabilidad.

²⁰ Nicola Lacey escribió la biografía de Hart en 2004, si se desea conocer más datos sobre la vida y bibliografía de este gran filósofo consúltese: Nicola Lacey, *A life of H. L. A. Hart. The nightmare and the noble dream*, Oxford University Press, Oxford, segunda reimpresión, 2012. Nicola Lacey también es abogada litigante y especialista en responsabilidad penal, participó en *The H.L.A. Hart Memorial Lecture* (2015) con el tema “Responsibility Without Consciousness” consúltese su texto en: *Oxford Journal of Legal Studies*, Volume 36, Issue 2, Summer 2016, Pages 219–241, <https://doi.org/10.1093/ojls/gqv032> <https://academic.oup.com/ojls/article/36/2/219/2472470> Sin entrar a su consulta, se recomienda también revisar: “Responsibility and the Negligence”, publicado en *Standard*, *Oxford Journal of Legal Studies*, Volume 30, Issue 1, Spring 2010, Pages 1–18, <https://doi.org/10.1093/ojls/gqq002> y Finnis, John “Equality and Differences”, *Solidarity: The Journal*

centra la discusión en la fórmula clásica del Derecho anglosajón *mens rea* y las condiciones necesarias para la responsabilidad penal (actos de voluntad, intención y negligencia). A la vez, hace reflexiones sobre si el Derecho Penal podría o debería prescindir de las condiciones mentales para asignar la responsabilidad. En el Derecho anglosajón existe la máxima *mens rea* (mente culpable), que se refiere al estado mental requerido por la ley para condenar a un acusado por un delito en específico (*actus non facit reum nisi mens sit rea*)²¹

Por lo general, en el Derecho anglosajón, es necesario establecer la máxima *mens rea* de un delincuente para probar la culpabilidad en un juicio penal. La acusación debe probar más allá de toda duda razonable que el acusado cometió el delito con un estado de ánimo culpable. El requisito de *mens rea* se basa en la idea de que uno debe poseer un estado mental culpable y ser consciente de su mala conducta; sin embargo, un acusado no necesita saber que su conducta es ilegal para ser culpable de un crimen. El imputado debe ser consciente de los hechos que hacen que su conducta se ajuste a la definición del delito.²²

Los estados mentales suelen estar organizados jerárquicamente por el estado de culpabilidad del delincuente. Generalmente, la culpabilidad del estado mental de un actor corresponde a la gravedad del delito. Los niveles más altos de culpabilidad normalmente se correlacionan con una responsabilidad más severa y sentencias más severas. Existe una clasificación tradicional o clásica de la doctrina penal contenida en el cuatro niveles: 1. Actuar intencionalmente: el acusado tenía un objeto consciente subyacente para actuar; 2. Actuar a sabiendas: el acusado está prácticamente seguro de que la conducta causará un resultado particular; 3. Actuar imprudentemente: el acusado ignoró conscientemente un riesgo sustancial e injustificado; y 4. Actuar con negligencia: el acusado no era consciente del riesgo, pero debería haberlo sabido.²³ En consecuencia, un delito cometido a propósito

of *Catholic Social Thought and Secular Ethics*, Vol. 2, número 1, disponible en: <https://researchonline.nd.edu.au/solidarity/vol2/iss1/1> (15 de junio de 2022).

²¹ Significa: “el acto no hace que la persona sea culpable a menos que la mente también sea culpable”. Véase, por ejemplo, *Staples vs. Estados Unidos*, 511 US 600 (1994)

²² Consúltase la voz *Mens rea*, Legal Information Institute, Cornell Law School en: https://www.law.cornell.edu/wex/mens_rea (15 de junio de 2022).

²³ *Idem*.

conllevaría un castigo más severo que si el delincuente actuara a sabiendas, imprudentemente o con negligencia.²⁴

Adicionalmente, existe lo que se llama delitos de responsabilidad objetiva. En este tipo de delitos, es suficiente que la fiscalía demuestre que el acusado cometió el acto ilícito, independientemente del estado mental del acusado. Por lo tanto, un estado mental de culpabilidad es irrelevante para un delito de responsabilidad objetiva. Ejemplos de delitos de responsabilidad objetiva en el Derecho Penal a menudo incluyen posesión de armas o estupro.²⁵

Hart comenta que, en todos los sistemas jurídicos avanzados, la responsabilidad de la condena por delitos graves se hace dependiente, no sólo de que el delincuente haya realizado los actos externos que la ley prohíbe, sino de que los haya realizado en un cierto estado de ánimo o con cierta voluntad. Las condiciones mentales o elementos mentales en la responsabilidad penal y, a pesar de mucha variación en los detalles y la terminología, son muy similares en la mayoría de los sistemas legales.

Lo que sí ha ocurrido es una disminución sobre el uso del *mens rea* (mente culpable) en la etapa de individualización de la pena (aunque se mantiene como un elemento necesario de la responsabilidad que debe establecerse antes de un veredicto). Hart comenta que no es algo que deba tenerse en cuenta para determinar la condena o disposición del condenado, aunque también puede considerarse para ese propósito.²⁶

En México, por ejemplo, el uso del *mens rea* en el tema de la individualización de la pena tiene limitantes con base en la Teoría del Derecho Penal del Acto que la Suprema Corte de Justicia de la Nación ha determinado se utilice por los jueces en esta etapa del procedimiento penal; y se basa en reconocer que nuestra constitución a partir de la reforma de justicia penal y seguridad pública de junio de 2008 y la reforma constitucional de Derechos Humanos de junio de 2011 adjudicó la Teoría

²⁴ Jules L. Coleman, The Morality of Strict Tort Liability, 18 Wm. & Mary L. Rev. 259, 1976, <https://scholarship.law.wm.edu/wmlr/vol18/iss2/3> (15 de junio de 2022).

²⁵ Consúltase la voz "Mens rea", *Legal Information Institute*, Cornell Law School en: https://www.law.cornell.edu/wex/mens_rea (15 de junio de 2022).

²⁶ *Punishment and Responsibility*, Clarendon Press, Oxford, 1968. p. 187.

del Derecho Penal del Acto (eliminando la Teoría del Derecho Penal del Actor). La Corte mexicana señala:

“...el derecho penal no puede sancionar la ausencia de determinadas cualidades o la personalidad, porque está limitado a juzgar actos... el derecho penal únicamente puede prohibir la comisión de conductas específicas (no la personalidad); es decir, sólo aquel acto prohibido por una norma penal, clara y explícita puede dar lugar a una sanción... El hecho de que la Constitución haya eliminado la posibilidad de que el sistema penal opere bajo la premisa de que alguien es desadaptado, fundamenta la convicción de que nuestro sistema se decanta por un derecho penal sancionador de delitos, no de personalidades. Así, el abandono del término “delincuente” también exhibe la intención del constituyente permanente de eliminar cualquier vestigio de un “derecho penal de autor”, permisivo de la estigmatización de quien ha cometido un delito. Esta conclusión se enlaza con la prohibición de penas inusitadas contenida en el artículo 22, primer párrafo, constitucional, la cual reafirma la prohibición de que cualquier consideración vinculada con etiquetas a la personalidad tenga incidencia en la punición.”²⁷

Mención especial merece el 400-Bis del Código Penal Federal (Operaciones con recursos de procedencia ilícita o lavado de activos), en el que el tipo penal pide que el sujeto tenga conocimiento de que los recursos proceden o representan el producto de una actividad ilícita. También ocurre lo mismo con el delito de Espionaje (129), Terrorismo (y sus diferentes modalidades contenidas en los artículos 139 Bis, 139 Quáter, 148 Ter); Sabotaje (140); Abuso de autoridad (215 fracciones VII y XI)

²⁷ La Primera Sala de la Suprema Corte de Justicia de la Nación, en las tesis jurisprudenciales 1a./J. 19/2014 (10a.), de rubro: “DERECHO PENAL DEL AUTOR Y DERECHO PENAL DEL ACTO. RASGOS CARACTERIZADORES Y DIFERENCIAS.” y 1a./J. 21/2014 (10a.), de rubro: “DERECHO PENAL DE ACTO. RAZONES POR LAS CUALES LA CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS SE DECANTA POR DICHO PARADIGMA (INTERPRETACIÓN SISTEMÁTICA DE LOS ARTÍCULOS 10., 14, TERCER PÁRRAFO, 18, SEGUNDO PÁRRAFO, Y 22, PRIMER PÁRRAFO).”, explicó las razones por las cuales -a partir de una interpretación sistemática de diversos principios constitucionales- es posible concluir que cuando los juzgadores se encuentran en condiciones de hacer uso de su prudente arbitrio para individualizar la pena, deben rechazar la posibilidad de ponderar la supuesta peligrosidad de la persona, así como cualquier prejuicio sobre alguna supuesta proclividad al delito, bajo la idea de que la persona cuenta con antecedentes penales. Las personas solamente pueden ser sancionadas por la comisión de conductas penales establecidas previamente en la ley; *nunca con apoyo en juicios de valor sobre su personalidad*. Sin embargo, el término “antecedentes penales”, entendido en sentido amplio -y que aplica para esta clase de valoraciones constitucionalmente vedadas- debe distinguirse del concepto de “reincidencia”, mismo que el legislador puede utilizar expresamente como criterio para elevar el parámetro de punibilidad de acuerdo con la jurisprudencia 1a./J. 80/2013 (10a.), de rubro: “INDIVIDUALIZACIÓN DE LA PENA. LOS ANTECEDENTES PENALES DEL SENTENCIADO QUE LLEVEN A CONSIDERARLO COMO REINCIDENTE, DEBEN TOMARSE EN CUENTA PARA FIJAR LA PUNIBILIDAD.”.

y muchos relacionados con el tema del uso o abuso del poder o las facultades para un fin que no fue conferido (corrupción) de muchas legislaciones.

Hart cita el caso de *The Director of Public Prosecutions v. Smith*²⁸ y la del juez Oliver Wendell Holmes en su libro *The Common Law*:

“Malice aforethought means any one or more of the following states of mind. (a.) An intention to cause the death of, or grievous bodily harm to, any person, whether such person is the person actually killed or not; (b.) Knowledge that the act which causes death will probably cause the death of, or grievous bodily harm to, some person, whether such person is the person actually killed or not, although such knowledge is accompanied by indifference whether death or grievous bodily harm is caused or not, or by a wish that it may not be caused; (c.) An intent to commit any felony whatever; (d.) An intent to oppose by force any officer of justice on his way to, in, or returning from the execution of the duty of arresting, keeping in custody, or imprisoning any person whom he is lawfully entitled to arrest, keep in custody, or imprison, or the duty of keeping the peace or dispersing an unlawful assembly, provided that the offender has notice that the person killed is such an officer so employed.”²⁹

Aquí la intencionalidad se mantiene presente y nos lleva a la conciencia y a la culpabilidad por el mismo conocimiento o la intención de hacer lo que la ley prohíbe. Hart comenta, adicionalmente, que las expresiones responsabilidad, responsable y responsable de abarcan una amplia gama de ideas diferentes, aunque se utilicen de manera estándar y pone un ejemplo con una historia de un capitán de barco borracho que perdió su barco en el mar:

²⁸ Cfr. The Law Commission, Imputed Criminal Intent (*Director of Public Prosecutions v. Smith*), Londres, Her Majesty's Stationery Office en: <https://s3-eu-west-2.amazonaws.com/lawcom-production-11jsxou24uy7q/uploads/2016/08/LC.-010-IMPUTED-CRIMINAL-INTENT-DIRECTOR-OF-PUBLIC-PROSECUTIONS-v.-SMITH.pdf>

²⁹ “Maldad de antemano significa uno o más de los siguientes estados mentales ..” (a.) La intención de causar la muerte o daño corporal grave a cualquier persona, ya sea que esa persona sea la persona realmente asesinada o no; (b.) El conocimiento de que el acto que causa la muerte probablemente causará la muerte o lesiones corporales graves a alguna persona, ya sea que esa persona sea realmente la persona muerta o no, aunque tal conocimiento vaya acompañado de indiferencia, ya sea de muerte o grave el daño corporal es causado o no, o por el deseo de que no sea causado; (c.) La intención de cometer cualquier delito grave; (d.) La intención de oponerse por la fuerza a cualquier funcionario de justicia en su camino hacia, en o al regresar de la ejecución del deber de arrestar, mantener bajo custodia o encarcelar a cualquier persona a quien legalmente tenga derecho a arrestar, mantener bajo custodia, o encarcelamiento, o el deber de mantener la paz o dispersar una reunión ilegal, siempre que el delincuente tenga conocimiento de que la persona asesinada es un oficial así empleado”. Cfr. Oliver Wendell Holmes, Jr., *The Common Law* en https://www.gutenberg.org/files/2449/2449-h/2449-h.htm#link2H_4_0002 Consúltase a Holmes para conocer un estudio histórico-comparado muy interesante sobre la evolución del concepto responsabilidad (capítulo primero de su libro).

Como capitán del barco, X era responsable de la seguridad de los pasajeros y la tripulación. Pero en su último viaje se emborrachaba todas las noches y era responsable de la pérdida del barco con todos a bordo. Se rumoreaba que estaba loco, pero los médicos consideraron que él era el responsable de sus acciones. Durante todo el viaje se comportó de manera bastante irresponsable, y varios incidentes en su carrera demostraron que no era una persona responsable. Siempre sostuvo que las excepcionales tormentas invernales fueron responsables de la pérdida del barco, pero en el proceso judicial entablado en su contra fue declarado responsable penalmente por su conducta negligente, y en procesos civiles separados fue considerado legalmente responsable por la pérdida de la vida y propiedad. Todavía está vivo y es moralmente responsable de la muerte de muchas mujeres y niños.³⁰

Este ejemplo, permite a Hart hacer un ejercicio de análisis y clasificación de la responsabilidad de cuatro tipos³¹:

- a. Responsabilidad como capacidad: Aquí Hart utiliza la frase: “Se rumoreaba que estaba loco, pero los médicos consideraron que él era el responsable de sus acciones.”, para ejemplificar la responsabilidad como capacidad, al afirmarse que la persona tiene “aptitudes normales” en su conducta. Estas capacidades son tomadas en cuenta por los sistemas jurídicos para la sancionabilidad y para la validez de los actos jurídicos. Las capacidades que se pueden incluir son “aquellas de entendimiento, razonamiento y control de la conducta.”³²
- b. Responsabilidad como factor causal: Aquí Hart utiliza la frase “Siempre sostuvo que las excepcionales tormentas invernales fueron responsables de la pérdida del barco” para ejemplificar la

³⁰ As captain of the ship, X was responsible for the safety of his passengers and crew. But on his last voyage he got drunk every night and was responsible for the loss of the ship with all aboard. It was rumoured that he was insane, but the doctors considered that he was responsible for his actions. Throughout the voyage he behaved quite irresponsibly, and various incidents in his career showed that he was not a responsible person. He always maintained that the exceptional winter storms were responsible for the loss of the ship, but in the legal proceedings brought against him he was found criminally responsible for his negligent conduct, and in separate civil proceedings he was held legally responsible for the loss of life and property. He is still alive and he is morally responsible for the deaths of many women and children. Cfr. *Punishment and Responsibility*, Clarendon Press, Oxford, 1968. p. 211.

³¹ La explicación de esta sección se hace con base en las ideas de Pablo Larrañaga Monjaraz, *El concepto de responsabilidad en la teoría del derecho contemporánea*, Tesis para obtener el grado de doctor en Derecho, Director: Manuel Atienza, Universidad de Alicante, Alicante, 1996, pp. 126-135.

³² *Idem*.

responsabilidad como un factor causal; es decir, consecuencias o resultados producidos no por la persona sino por otra persona, un animal, alguna situación o evento de la naturaleza. Se dice que “este uso sólo se da para referirse a relaciones causales cuyo resultado es desafortunado o grato... la relación causal entre una acción y un resultado ilícito es una de las condiciones necesarias para hablar de responsabilidad como sancionabilidad.”.³³

- c. Responsabilidad como deberes y obligaciones derivados de un cargo o rol social: Aquí Hart utiliza la frase “Como capitán del barco, X era responsable de la seguridad de los pasajeros y la tripulación” para ejemplificar la responsabilidad como deberes y obligaciones derivados de un cargo o rol social. El término responsable se usa para “designar un conjunto de deberes y obligaciones que tenía X por el hecho de ser el capitán del barco... podemos referirnos tanto a deberes jurídicos como deberes morales”³⁴.
- d. Responsabilidad como sancionabilidad: Aquí Hart utiliza la frase “en el proceso judicial entablado en su contra fue declarado responsable penalmente por su conducta negligente, y en procesos civiles separados fue considerado legalmente responsable por la pérdida de la vida. y propiedad” para explicar la responsabilidad como sanción. “Si bien las expresiones generales “responsabilidad” y “sancionabilidad” (*liability*) son expresiones equivalentes en muchos contextos, el significado de la afirmación de que una persona es responsable de algo, o por algún acto o por algún daño no es siempre idéntico al significado de la afirmación de que una persona es sancionable por el acto o del daño causado... La responsabilidad en este sentido parece unificar los otros sentidos que venimos analizando”³⁵.

³³ *Idem.*

³⁴ *Idem.*

³⁵ *Idem.* Para conocer más sobre los tipos de responsabilidad de Hart, consúltese: *Punishment and Responsibility*, Oxford, Clarendon Press, 1968. pp. 211 y ss.

Ronald Dworkin. Para concluir, el último filósofo que expondremos es Ronald Dworkin (1931-2013), basándonos en su libro *Justicia para erizos*. Dworkin explica los vínculos de la responsabilidad moral y la responsabilidad jurídica y la relevancia de reflexionar sobre estos temas. Hace una propuesta de ver a la responsabilidad moral como una virtud que se debe fomentar y cultivar en los ciudadanos, nos movemos al ámbito de la responsabilidad jurídica con soporte de la moral. Dworkin explica que el carácter y la magnitud de nuestra responsabilidad por las acciones que emprendemos giran más bien alrededor de una cuestión ética: ¿cuál es la índole de una vida bien vivida? ¿Cómo deberíamos vivir? “cada uno de nosotros tiene la obligación Ética soberana de hacer algo de valor de su propia vida... la ambición de hacer de nuestra vida una buena vida: auténtica y valiosa... En especial, debemos atesorar nuestra dignidad. El concepto de dignidad se ha desvalorizado a causa de su uso inconsistente y excesivo en la retórica política... Pero necesitamos la idea, y la idea afín de autorrespeto... me apoyo en la tesis de Immanuel Kant de que no podemos respetar adecuadamente nuestra humanidad a menos que respetemos la humanidad de otros.”³⁶

Para Dworkin, cualquier responsabilidad moral debe de pasar el filtro del libre albedrío y actualiza el debate para recordarnos las posturas filosóficas sobre el tema: “Las personas son responsables de sus actos solo cuando controlan lo que hacen; sólo... cuando tienen libre albedrío y actúan movidas por él.”³⁷ Y, a la par, cita la postura relativa a la inexistencia del libre albedrío: “En realidad, la gente nunca controla su comportamiento ni siquiera cuando parece que lo hace. Su voluntad nunca es libre porque su comportamiento siempre tiene por causa alguna combinación de fuerzas y sucesos que están enteramente al margen del control vigente en su cerebro. Nunca es cierto que podrían haber hecho otra cosa y no lo que hicieron. En rigor, estas decisiones de la gente no solo son causadas por sucesos previos, sino que ni siquiera causan las acciones de las cuales ella se cree

³⁶ Ronald Dworkin, *Justicia para erizos*, México, FCE, 2014, p. 26-31.

³⁷ *Ibidem*. p. 272.

responsable. En consecuencia, la responsabilidad es una ilusión, y siempre es inapropiado considerar a las personas culpables o castigarlas por lo que hacen.”.³⁸

Algunos filósofos afirman que, si esta tesis es verdadera, entonces el concepto de responsabilidad moral carecería de aplicación práctica al momento de que se castiga o se sanciona conforme a lo que dice la norma jurídica. Thomas Nagel³⁹ y Peter Strawson⁴⁰ son utilizados por Dworkin para explicar este tema y, los problemas que se presentan del llamado desafío del libre albedrío, el cual está lleno de discusiones relacionadas con las causas y consecuencias del pensamiento y la acción.⁴¹ Otros filósofos enfocan sus discusiones hacia la libertad y discusiones respecto la responsabilidad en juicio.

Dworkin se pregunta: ¿Está todo el comportamiento humano completamente determinado por acontecimientos previos sobre los cuales las personas mismas no ejercen ningún control? ¿Hay algún comportamiento humano que sea causado por acontecimientos físicos o biológicos aleatorios y casuales sobre los cuales la gente tampoco tenga control?⁴²

El sistema de responsabilidad -afirma Dworkin- incorpora principios éticos que conectan las causas de nuestras decisiones con nuestra responsabilidad. El sistema de responsabilidad muestra que solo somos responsables cuando estamos en control de nuestro comportamiento.⁴³ Este es un ingrediente medular, que se debe tomar en cuenta para aquellas personas que crean, utilizan o deciden a través

³⁸ *Idem*. Dworkin define: a) “decisión” para describir el conocido suceso consciente que percibimos como decidir e incluye decisiones reflexivas meditadas que tomamos al cabo de una deliberación y decisiones irreflexivas que tomamos minuto a minuto para seguir con lo que estábamos haciendo en vez de hacer otra cosa; b) “determinismo” todas esas decisiones, tanto reflexivas como irreflexivas, están íntegramente determinadas por procesos y sucesos que las preceden y están al margen del control del decisor; c) “epifenomenalismo”, niega que las decisiones aparezcan siquiera en la cadena causal que termina en movimientos de los nervios y los músculos. Supone que la sensación interna de haber decidido hacer algo es solo un efecto secundario de los sucesos físicos y biológicos que realmente produjeron el comportamiento decidido; d) “voluntad”, a la parte de mí que toma la decisión. Tanto el determinismo como el epifenomenalismo siguen sin ser demostrados. Todo es posible hasta ahora. Cfr. Ronald Dworkin, *op. cit.*, pp. 272-273.

³⁹ Thomas Nagel, *The View from Nowhere*, Oxford, Oxford University Press, 1986, p. 111. Existe una traducción al español del Fondo de Cultura Económica (1996)

⁴⁰ Peter Strawson “Freedom and Resentment” en *Freedom and Resentment, and Other Essays*, Londres, Methuen, 1974. Existe una versión en español de editorial Paidós, 1995.

⁴¹ Cfr. Ronald Dworkin, *op. cit.*, p. 274.

⁴² Ronald Dworkin, *op. cit.*, p. 276.

⁴³ *Idem*.

de sistemas inteligentes en los que su voluntad, su conciencia o su comportamiento no se ven reflejados ni en el proceso de entrenamiento ni en el proceso de uso o aplicación y, principalmente, cuando utilizamos redes neuronales convolucionales profundas para nuestro modelo de procesamiento de datos y sus resultados.

Solo cuando nuestra voluntad y conciencia están a cargo/o al mando, podemos dar o negar valor a nuestra responsabilidad. El dilema del tranvía⁴⁴ puede funcionar bien para explicar la responsabilidad de una persona que conduce un vehículo automotor o que toma una decisión en la que va implícita una afectación a otro u otros. Sin embargo, la decisión la sigue tomando y está bajo la voluntad, conciencia y responsabilidad de la persona; pero no funciona en un modelo en el que un humano no controla el vehículo ni sus decisiones algorítmicas, decidiendo el sistema la trayectoria sin su participación. La responsabilidad de la persona -física o moral- no es tan clara al momento de no tener el control.

Dworkin continúa diciendo que hay una interpretación alternativa del significado de tener el control. Según este otro punto de vista “un agente tiene el control cuando es consciente de enfrentar y tomar una decisión, cuando ninguna otra toma esa decisión por su intermedio y para él y cuando el propio agente tiene las capacidades para forjarse creencias verdaderas sobre el mundo y hacer coincidir sus decisiones con su personalidad normativa: sus deseos, ambiciones y convicciones establecidas. Este es el sentido del control basado en la capacidad.”.⁴⁵

El control basado en capacidades de las personas es el que mejor sirve y puede utilizarse para proteger a la empresa de conductas relacionadas con el cumplimiento normativo, tanto interno como externo. En el caso de la responsabilidad penal de las personas jurídicas⁴⁶ o la responsabilidad

⁴⁴ Jaime Rubio Hancock, *El dilema del tranvía: ¿debo sacrificar una vida para salvar cinco?*, Verne, 31 de marzo de 2017, en: [¿Cómo funciona Adext AI? Explore la IA y amplifiquen | Centro de ayuda de Adext AI](#)

⁴⁵ *Idem*.

⁴⁶ Desde 2014, México incorpora a su legislación esta tendencia internacional derivada de la Convención de Mérida (2005). Se incorpora en el Código Penal Federal (Artículo 11 Bis) y en el Código Nacional de Procedimientos Penales (Título X, Capítulo II). En todos los supuestos previstos en el artículo 422 del Código Nacional de Procedimientos Penales -dice el referido artículo 11 Bis, en su último párrafo- las sanciones podrán atenuarse hasta en una cuarta parte, si con anterioridad al hecho que se les imputa, las personas jurídicas contaban con un órgano de control permanente, encargado de *verificar el cumplimiento de las disposiciones legales aplicables para darle seguimiento*

administrativa⁴⁷ (ambas contenidas en leyes mexicanas), el *Compliance* penal o administrativo, a su vez, aplica mecanismos de prevención de riesgos para detectar si el cuerpo directivo de la organización o sus trabajadores están cumpliendo con su responsabilidad de supervisión y de actuación y reporte de sus actividades. Los riesgos reputacionales, operativos y administrativos inician desde la etapa del reclutamiento del personal y permanecen durante todo el tiempo de vida de la empresa y durante todo el tiempo de permanencia de sus empleados.

Siguiendo a Dworkin, podemos decir que existe el “principio del control casual” que ve la cuestión de la responsabilidad al margen de la percepción corriente que un agente tiene de su situación; y, el “principio del control basado en la capacidad” que pone la responsabilidad entre “los paréntesis de una vida corriente vivida desde una perspectiva personal.”.⁴⁸

La responsabilidad se asocia a las decisiones finales. Por ello, la Ética en la empresa es más que la norma impuesta.⁴⁹ Para explicar esto, Dworkin pone el ejemplo del delito de homicidio y la tentativa de homicidio, donde los abogados

a las políticas de prevención delictiva y que hayan realizado antes o después del hecho que se les imputa, la disminución del daño provocado por el hecho típico.

⁴⁷ La Ley General de Responsabilidades Administrativas en su artículo 25 señala: En la determinación de la responsabilidad de las personas morales, se valorará si cuentan una política de integridad. Se considerará una política de integridad aquella que cuenta con, al menos, los siguientes elementos: I. Un manual de organización y procedimientos que sea claro y completo, en el que se delimiten las funciones y responsabilidades de cada una de sus áreas, y que especifique claramente las distintas cadenas de mando y de liderazgo en toda la estructura; II. Un código de conducta debidamente publicado y socializado entre todos los miembros de la organización, que cuente con sistemas y mecanismos de aplicación real; III. Sistemas adecuados y eficaces de control, vigilancia y auditoría, que examinen de manera constante y periódica el cumplimiento de los estándares de integridad en toda la organización; IV. Sistemas adecuados de denuncia, tanto al interior de la organización como hacia las autoridades competentes, así como procesos y consecuencias concretas respecto de quienes actúan de forma contraria a las normas internas o a la legislación mexicana; V. Sistemas y procesos adecuados de entrenamiento y capacitación respecto de las medidas de integridad; VI. Políticas de recursos humanos tendientes a evitar la incorporación de personas que puedan generar un riesgo a la integridad de la corporación. y VII. Mecanismos que aseguren en todo momento la transparencia y publicidad de sus intereses.

⁴⁸ Para Dworkin, estos principios son antagónicos. La existencia de uno no es compatible con el otro, sin embargo, ambos son principios de carácter “ético o moral”, no físico, biológico o metafísico. Y también son principios interpretativos, no se deriva de ningún descubrimiento científico o metafísico.

⁴⁹ Dos excelentes libros sobre la Ética en la empresa que sirven de guía para su implementación son: Carlos Llano Cifuentes, *Dilemas éticos de la empresa contemporánea*, IPADE-Ediciones ECA, México, 2019 y Bernardo Fernández Ardaín, *Ética de la Dirección. Decisión con libertad*, México, IPADE-Ediciones Ruz, 2006.

descubren o inventan razones para explicar por qué la tentativa se debe sancionar con una menor penalidad que el homicidio.

Es importante señalar que Dworkin se pronuncia a favor del principio de capacidad ya que considera que abarca mucho mejor la totalidad de posturas Éticas y filosóficas.⁵⁰ Aquí las decisiones están determinadas y emanan del carácter. Las discapacidades y los accidentes son diferentes ya que no reflejan el carácter. El principio de capacidad “describe competencias que, a nuestro entender, una persona debe tener si pretendemos que sea razonable juzgarla exitosa o no exitosa en sus esfuerzos por vivir bien... El principio de capacidad sostiene... que una persona carece de control... cuando no tiene la capacidad suficiente para formarse creencias verdaderas y pertinentes sobre el mundo en el cual actúa o para hacer coincidir sus decisiones con su personalidad normativa.”.⁵¹

Gran parte del debate que existe entre abogados y no abogados es en realidad una discusión sobre dónde debería establecerse dicho umbral. Son disputas que giran alrededor de micro juicios de valor que las personas que aceptan el principio de capacidad en abstracto harán de diferente manera. Dworkin reconoce que una enfermedad mental puede desbaratar cualquiera de las dos capacidades de juicio y concluye:

“La elección entre los dos principios, el causal y el basado en la capacidad, es importante por razones que van mucho más allá de la polémica sobre el libre albedrío... esa decisión es decisiva para explicar y discutir las controversias de carácter mucho más práctico... entre personas que aceptan la estructura general del sistema de la responsabilidad, pero discrepan en torno de su aplicación a casos específicos. Si estimamos que las personas solo son responsables cuando sus acciones emanan de un acto espontáneo e incausado de la voluntad, consideraremos que esas controversias prácticas giran alrededor de un hecho psicobiológico planteado en términos de todo o nada.”.⁵²

El principio de causalidad y el basado en la capacidad sirven para categorizar los casos que se presentan en el Derecho y la responsabilidad de las personas; como en otras ocasiones, Dworkin está preocupado por dar soluciones a los operadores

⁵⁰ Aristóteles, Hobbes, Hume aceptaban el principio de capacidad.

⁵¹ *Ibidem.*, pp. 300 y 301.

⁵² *Ibidem.*, p. 308.

del Derecho utilizando fórmulas de corte moral, usando principios y dando soporte a ese ideal que debemos alcanzar como seres humanos: tener la libertad para buscar vivir bien y alcanzar nuestras metas. La libertad no reside en la conciencia del fin al que se dirige -diría Carlos Llano- sino en la determinación precisa del mismo fin al que quiere dirigirse. “Siendo, como es el fin, el elemento más importante de toda acción, la causa que domina a todas las demás causas -*causa causarum*-, ese acto de libertad por el que decidimos el fin de nuestra vida llega a decidir así, en cierto modo todo lo que somos”⁵³. Veamos ahora lo que la legislación civil y penal dicen sobre la responsabilidad, aquí analizaremos al hecho ilícito -como principal fuente de la responsabilidad civil- y su vinculación con los delitos en materia penal. Posteriormente, estudiaremos la importancia del concepto acto médico y las características particulares de la responsabilidad del médico en materia civil, penal y patrimonial del Estado.

⁵³ Carlos Llano Cifuentes, *Las formas actuales de la libertad*, México, Trillas, 2014, p. 108.

A) Categorías de la responsabilidad jurídica

En este apartado explicaremos la responsabilidad civil, penal y aquellas categorías de responsabilidad relacionadas específicamente con el actuar del médico (acto médico).

Categorías de responsabilidad civil. En el Derecho Civil mexicano, responsabilidad surge del hecho ilícito o el riesgo creado y tiene relación con dos categorías de la responsabilidad jurídica: responsabilidad subjetiva y responsabilidad objetiva. También encontramos una forma de ubicar la responsabilidad derivada de los actos contractuales o de los actos extracontractuales.

Rojina Villegas comenta que la responsabilidad subjetiva deriva de la Teoría de la Culpa (parte de la conducta del sujeto, dolo, negligencia o culpa) y la responsabilidad objetiva, en cambio, prescinde de este elemento interno o subjetivo, para tomar en cuenta una actividad en la que el ser humano utiliza alguna cosa peligrosa que causa daño (y una relación causa efecto entre el hecho y el daño)⁵⁴.

La responsabilidad subjetiva contempla una regla clásica “si el oferente se retracta, comete un hecho ilícito, y está obligado al pago de daños y perjuicios”⁵⁵ y cita el artículo 1910 del Código Civil Federal: *El que obrando ilícitamente o contra las buenas costumbres cause daño a otro, está obligado a repararlo, a menos que demuestre que el daño se produjo como consecuencia de culpa o negligencia inexcusable de la víctima.*

Respecto a la responsabilidad objetiva, Rojina Villegas ubica sus elementos: “1.- El uso de cosas peligrosas, 2.- La existencia de un daño de carácter patrimonial, 3.- La relación de causa a efecto entre el hecho y el daño”.⁵⁶ El ser humano, conforme a esta teoría, genera riesgos con determinados actos que le generan algún beneficio, por lo que debe hacerse responsable de estos daños. Al recibir un beneficio de esta actividad, “es justo que repare el daño que ocasiona... *Ubi*

⁵⁴ Rafael Rojina Villegas, *Compendio de Derecho Civil*, XXVI edición, Tomo III, México, Porrúa, 2006, p. 275 y 276.

⁵⁵ *Ibidem.*, p. 225.

⁵⁶ *Ibidem.*, p. 275.

emolumentum ibi onus, decían los romanos. A los provechos deben corresponder las pérdidas”.⁵⁷

Ernesto Gutierrez y González, comenta que el Código Penal de 1871 para el Distrito Federal y Territorios de Baja California, en su artículo 301 dispuso que la responsabilidad civil proveniente de un hecho u omisión contrarios a una ley penal, consiste en la obligación de que el responsable haga: 1.- La restitución; II.- La reparación; III.- La indemnización; y, IV.- El pago de gastos judiciales.⁵⁸

Responsabilidad por hecho ilícito -dice Gutiérrez y González- es una conducta que consiste en restituir las cosas al estado que tenían y de no ser posible, a la restitución del detrimento patrimonial (daño y/o perjuicio) generado por una acción u omisión de quien lo cometió por sí mismo, o esa acción u omisión permitió que se causara el detrimento, por personas a su cuidado, o cosas que posee, y que originó con ello la violación culpable de un deber jurídico *stricto sensu*, o de una obligación *lato sensu* previa, en cualquiera de sus dos especies.⁵⁹ Y se debe tener presente que surge por un hecho ilícito no penal.

Al analizar los elementos de la responsabilidad civil por hecho ilícito⁶⁰ explica lo siguiente:

1. Una acción o una omisión: El hecho ilícito deriva de una acción que consista en hacer lo contrario a lo que un deber jurídico determina o a lo establecido en una obligación previa, *lato sensu*; o bien una omisión, que consiste en no hacer lo que el deber jurídico, la obligación previa contractual o una declaración unilateral de voluntad manda.
2. Un detrimento patrimonial: El detrimento patrimonial se traduce en un daño o en un perjuicio. Se entiende por daño la pérdida o menoscabo sufrido en el patrimonio por la falta de cumplimiento de una obligación (Artículo 2108, Código Civil Federal). Se reputa perjuicio la privación de cualquiera ganancia

⁵⁷ *Ibidem.*, p. 276.

⁵⁸ Cfr. Gutiérrez y González, Ernesto, *Derecho de las obligaciones*, XXIII Edición, México, Porrúa, 2020, p. 560.

⁵⁹ Ernesto Gutierrez y González, *op. cit.*, p. 562.

⁶⁰ Artículo 1830.-Es ilícito el hecho que es contrario a las leyes de orden público o a las buenas costumbres.

lícita, que debiera haberse obtenido con el cumplimiento de la obligación (Artículo 2109, Código Civil Federal).

3. Relación de causalidad entre la conducta y el daño o perjuicio: Debe existir un vínculo inmediato y directo entre la conducta ilícita y el daño o perjuicio (lo mismo sucede para la responsabilidad objetiva).⁶¹
4. Restitución de las cosas al estado que tenían: El artículo 1915 del código en comento nos guía en este tema: La reparación del daño debe consistir a elección del ofendido en el restablecimiento de la situación anterior, cuando ello sea posible, o en el pago de daños y perjuicios. Cuando el daño se cause a las personas y produzca la muerte, incapacidad total permanente, parcial permanente, total temporal o parcial temporal, el grado de la reparación se determinará atendiendo a lo dispuesto por la Ley Federal del Trabajo. Para calcular la indemnización que corresponda se tomará como base la Unidad de Medida y Actualización y se extenderá al número de unidades que para cada una de las incapacidades mencionadas señala la Ley Federal del Trabajo. En caso de muerte la indemnización corresponderá a los herederos de la víctima. Los créditos por indemnización cuando la víctima fuere un asalariado son intransferibles y se cubrirán preferentemente en una sola exhibición, salvo convenio entre las partes. Las anteriores disposiciones se observarán en el caso del artículo 2647 del Código Civil Federal, es decir, respecto los portadores.⁶²

⁶¹ El artículo 2110 del Código Civil Federal dice: “Los daños y perjuicios deben ser consecuencia inmediata y directa de la falta de cumplimiento de la obligación, ya sea que se hayan causado o que necesariamente deban causarse.”. Existe una tesis aislada de tribunales colegiados que señala que en este artículo se ha querido, excluir del resarcimiento todos aquellos daños y perjuicios que no deriven directa e inmediatamente del evento dañoso, por ser a su vez producidos por alguno de los efectos del propio evento, quedando entonces limitada la responsabilidad a los primeros; en caso contrario no habría límite para la responsabilidad y el obligado tendría que pagar daños y perjuicios. Cfr. “DAÑOS Y PERJUICIOS. DEBEN SER CONSECUENCIA INMEDIATA Y DIRECTA DEL EVENTO DAÑOSO.” [TA]; 10a. Época; T.C.C.; Gaceta S.J.F.; Libro 60, noviembre de 2018; Tomo III; Pág. 2207 I.8o.C.68 C (10a.) Registro digital: 2018297

⁶² “Art. 2647.- Los portadores responden del daño causado a las personas por defecto de los conductores y medios de transporte que empleen; y este defecto se presume siempre que el empresario no pruebe que el mal aconteció por fuerza mayor o por caso fortuito que no le puede ser imputado.

5. De no ser posible la restitución, entonces el daño y el perjuicio se pagan con dinero: Es decir, si las cosas no se pueden volver al estado en que se encontraban antes de la acción o la omisión que las afectó, entonces se deben pagar en dinero.
6. Imputable al autor de la acción u omisión: El perjudicado por el hecho ilícito solo tiene que probar que la conducta dañosa le es imputable al autor del acto u omisión para que se genere la responsabilidad y la consecuente reparación.
7. Que la acción u omisión sea de un hecho propio o que, con su acción u omisión, origine que una persona a su cuidado, o cosas que posee, sean las que físicamente causen el daño material: La responsabilidad por el hecho ilícito puede recaer en quien verifica la acción o realiza la omisión; pero también, puede recaer la responsabilidad -en esa misma persona- aunque no participe físicamente al tener a su cuidado alguna persona o cosa que posea (siendo una u otra la generadora del hecho ilícito).⁶³

Existen casos de excepción en los que una persona produce una conducta con la que comete un hecho ilícito y se causa una afectación patrimonial que la ley exime de la obligación de repararla:

1. Caso fortuito o fuerza mayor. – Puede afectar cualquier tipo de responsabilidad. Es un “fenómeno de la naturaleza, o un hecho de persona con autoridad pública, temporal o definitivo, general -salvo caso excepcional, - insuperable, imprevisible, o que previéndose no se puede evitar, y que origina que una persona realice una conducta que produce a otra persona, un detrimento patrimonial, contraria a un deber jurídico stricto sensu o a una obligación lato sensu.”⁶⁴

⁶³ Para la elaboración de las ideas de este apartado se tomó como guía la “Teoría unitaria de la responsabilidad jurídica” de Ernesto Gutiérrez y González.

⁶⁴ Cfr. Gutiérrez y González, *op. cit.*, pp. 606 y ss. En la legislación mexicana, dice Gutiérrez y González, estos conceptos son sinónimos y el Código Civil Federal los emplea indistintamente. Sin embargo, en otros países como Francia, si hay diferencias. Caso fortuito se refiere a los acontecimientos de la naturaleza; y fuerza mayor, a hechos del hombre. El artículo 2111 del Código Civil Federal contiene una excepción para el supuesto de caso fortuito o fuerza mayor al señalar que “nadie está obligado al caso sino cuando ha dado causa contribuido a él, cuando ha aceptado expresamente esa responsabilidad, o cuando la ley se la impone.” Es decir, si el autor de la conducta dañosa contribuyó al incumplimiento de su deber u obligación o cuando expresamente se acepta

2. Cuando se produce como consecuencia de culpa o negligencia “inexcusable” de la víctima. - Este supuesto se encuentra en la última parte del artículo 1910 del Código Civil Federal al mencionar aquellos casos en donde el daño se deba a culpa o negligencia inexcusable de la víctima, de tal forma que “quien materialmente produce el daño, queda relevado de la responsabilidad de indemnizar.”.⁶⁵
3. Cuando cada una de las partes soportan lo ocurrido sin derecho a indemnización.- Este lo encontramos en el artículo 1914 del Código Civil Federal cuando señala que “...cuando sin el empleo de mecanismos, instrumentos, etc., ... y sin culpa o negligencia de ninguna de las partes se producen daños, cada una de ellas los soportará sin derecho a indemnización.”.
4. Prescripción. - Es “la facultad o derecho que la ley establece a favor del deudor, para que se excepcione válidamente y sin responsabilidad, de cumplir con la prestación debida, o para exigir ante la autoridad competente la declaración de que ya no se le puede exigir en forma coactiva la prestación debida, cuando ha transcurrido el plazo que otorga la ley a su acreedor para hacer efectivo su derecho. El Código Civil Federal, regula esta forma de no responsabilidad cuando dice: “Prescriben en dos años: ...V. La responsabilidad civil proveniente de actos ilícitos que no constituyan delitos.”.

La responsabilidad civil por hecho ilícito corresponde eminentemente del ámbito del Derecho Civil y no del ámbito penal. La responsabilidad penal, a diferencia de la responsabilidad civil, genera una sanción pecuniaria que cubrirá el

responder al daño aunque se hubiera causado por caso fortuito o fuerza mayor o cuando la propia ley lo impone, por ejemplo, en el 812-II y 2505. “Artículo 812.-El que posee por menos de un año, a título traslativo de dominio y con mala fe, siempre que no haya obtenido la posesión por un medio delictuoso, está obligado: ...II. A responder de la pérdida o deterioro de la cosa sobrevenidos por su culpa, o por caso fortuito o fuerza mayor, a no ser que pruebe que éstos se habrían causado aunque la cosa hubiere estado poseída por su dueño. No responde de la pérdida sobrevenida natural o inevitablemente por el sólo transcurso del tiempo. Tiene derecho a que se le reembolsen los gastos necesarios... Artículo 2505.-Si la cosa perece por caso fortuito, de que el comodatario haya podido garantizarla empleando la suya propia, o si no pudiendo conservar más que una de las dos, ha preferido la suya, responde de la pérdida de la otra.”.

⁶⁵ “Artículo 1910.-El que obrando ilícitamente o contra las buenas costumbres cause daño a otro, está obligado a repararlo, a menos que demuestre que el daño se produjo como consecuencia de culpa o negligencia inexcusable de la víctima.”

sentenciado y ésta comprende una multa⁶⁶ y la reparación del daño⁶⁷. La reparación del daño proveniente del delito que tiene “carácter de pena pública”⁶⁸. La reclamación de la reparación del daño, que no se obtenga por la vía penal ya sea por no ejercicio de la acción o la abstención de investigar por parte del ministerio público, sobreseimiento o sentencia absolutoria “podrá recurrirse por la vía civil o administrativa”⁶⁹. Y, cuando la reparación del daño proveniente de un delito provenga de un tercero, tendrá el carácter de responsabilidad civil.⁷⁰ En palabras de Rojina Villegas, cuando un hecho ilícito causa daño, pero no encuadra en un tipo penal, es entonces, un hecho ilícito civil. Salvo y no está sancionado por el Código Penal, acción u omisión no encuadra en un tipo penal, pudiera convertirse en un hecho ilícito y también habrá hechos ilícitos “que no son ni delitos penales ni civiles, porque no tienen sanción pública y porque no originan daño de carácter patrimonial o moral; por consiguiente, no son fuente de obligaciones desde el punto de vista civil.”⁷¹

La Primera Sala de la Suprema Corte de Justicia de la Nación, en una tesis de 2018, explica cómo se debe resolver un asunto de reparación del daño por hecho ilícito (sea civil extracontractual o administrativa por actividad irregular del Estado) y la necesidad de acreditar caso por caso: 1) la existencia del hecho ilícito, 2) la materialización del daño, y 3) la existencia de un nexo causal entre ambos. Lo que debe cuidar el juzgador en todo momento, dice la Primera Sala, es:

“...primero, que las normas y los procedimientos en que se sustente cada uno de los elementos descritos sean válidos a la luz del parámetro de control de regularidad; segundo, que la noción de ilicitud sea compatible con los estándares de Derechos Humanos que eventualmente resulten aplicables, partiendo de la posible existencia de Derechos Humanos subyacentes a las relaciones jurídicas que se estudien; y tercero, que la reparación que en su caso se dicte sea compatible con los estándares de reparación integral del daño o de justa indemnización. En relación con este último punto, las materias civil y administrativa cuentan también con una serie de reglas y principios que rigen la cuantificación de las indemnizaciones y la individualización de las medidas de reparación que puedan dictarse. Así, en términos de los artículos 10. de la

⁶⁶ Ver artículo 29 del Código Penal Federal mexicano.

⁶⁷ Ver artículo 30 del Código Penal Federal mexicano.

⁶⁸ Ver artículo 34 del Código Penal Federal mexicano.

⁶⁹ *Idem.*

⁷⁰ *Idem.*

⁷¹ Cfr. Rafael Rojina Villegas, op. cit., pp. 290 y 291.

Constitución Política de los Estados Unidos Mexicanos y 63.1 de la Convención Americana sobre Derechos Humanos, lo importante será que las reglas previstas en cada materia permitan que las indemnizaciones que resulten procedentes, sean compatibles con el derecho a una justa indemnización, atendiendo a la naturaleza del procedimiento en que se actúa.”⁷²

Al lado de la categoría de responsabilidad por hecho ilícito está otra categoría de responsabilidad que surge con la revolución industrial y el maquinismo o responsabilidad objetiva (por riesgo creado). Esta responsabilidad derivada de una conducta o un hecho previsto en la ley como objetivamente dañoso derivado del uso de un objeto que la ley considera por sí mismo peligroso o derivado de una conducta errónea hecha de buena fe.⁷³ Hay que decir que los códigos penales también regulan el uso de vehículos o maquinaria y el daño causado.⁷⁴

A decir de Gutiérrez y González, la responsabilidad objetiva no tiene un origen derivado del Derecho Penal, “se gesta en Roma, y se presenta en varios casos no sistematizados, pero ahí está, y luego en el Código Napoleón, y de ahí pasó a todos los Códigos modernos... La responsabilidad objetiva en su especie de riesgo creado, y sí, ahí es en donde se elabora una teoría parcial de la responsabilidad objetiva, ya que se la refiere sólo al riesgo creado, y se olvidan todos los demás casos en que surge esta responsabilidad sin culpa.”⁷⁵

⁷² “DERECHO A UNA JUSTA INDEMNIZACIÓN POR VIOLACIONES A DERECHOS HUMANOS. SU RELACIÓN CON EL DERECHO DE DAÑOS”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 293 1a. CLXXXIX/2018 (10a.) Registro digital: 2018646. Llama la atención, además de este criterio novedoso, la recomendación de la SCJN para lograr una indemnización justa en casos que involucren al Estado o a sus agentes como parte demandada: “A. Factores para ponderar respecto a la víctima, que incluyen (a) Aspecto cualitativo del daño o daño moral en sentido estricto, el cual comprende la valoración de (i) el tipo de derecho o interés lesionado, (ii) la existencia del daño, y (iii) la gravedad de la lesión o daño; y (b) Aspecto patrimonial o cuantitativo derivado del daño moral, dentro del cual se deberán valorar (i) los gastos devengados derivados del daño moral, y (ii) los gastos por devengar. B. Factores a ponderar respecto de los sujetos responsables: (i) naturaleza de la relación jurídica en el marco de la cual tuvo lugar el hecho ilícito; (ii) grado de responsabilidad; (iii) capacidad económica; y (iv) finalidad y objetivo de la indemnización (reparar, prevenir o reprimir).” “DERECHO A UNA JUSTA INDEMNIZACIÓN. DETERMINACIÓN DEL QUÁNTUM EN LOS CASOS EN QUE EL ESTADO ES LA PARTE DEMANDADA.” [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 289 1a. CXCIII/2018 (10a.) Registro digital: 2018643

⁷³ Gutiérrez y González, *op. cit.*, pp. 794-795.

⁷⁴ Un ejemplo es el artículo 172: “Cuando se cause algún daño por medio de cualquier vehículo, motor o maquinaria, además de aplicar las sanciones por el delito que resulte, se inhabilitará al delincuente para manejar aquellos aparatos, por un tiempo que no baje de un mes ni exceda de un año. En caso de reincidencia, la inhabilitación será definitiva.”

⁷⁵ *Idem.*

Existe una tesis de los Tribunales Colegiados de Circuito, del año 1996, que analiza este tema de manera singular y se intitula: RESPONSABILIDAD CIVIL POR HECHOS PROPIO, AQUILIANA Y OBJETIVA. DIFERENCIAS. La tesis dice:

“Los hechos lícitos y los ilícitos generan obligaciones; y así, es regla que la conducta de una persona le es imputable a ella; por esto, a la responsabilidad proveniente de la conducta de una persona, sea que esa conducta sea lícita o ilícita, se le llama subjetiva porque implica el elemento culpa. Como excepción a dicha regla, se establece que la conducta de terceros también sea imputable a otras personas, a ésta se le llama responsabilidad aquiliana en razón del jurisconsulto romano que creó la fórmula; en esta figura el elemento culpa se encuentra desvanecido, porque se reconoce que la conducta que causó un daño, es ajena a quien resulta obligado, pero aun así, se estima que tiene una culpa por falta de cuidado en las personas que de él dependen y cuya conducta causara el daño, que a su vez, generara una obligación, no a quien lo cometió, sino a la persona de quien dependiera. Por ello, incurren en tal responsabilidad los padres respecto de sus hijos, los mentores respecto de sus pupilos dentro del recinto educativo, los hoteleros respecto de sus empleados, los patrones respecto de sus trabajadores y el Estado respecto de sus servidores. Diversa excepción es la que resulta aun ante la ausencia de conducta, por el solo hecho de ser dueño de una cosa que por sí misma causa un daño. Aquí, no hay conducta y por lo mismo no hay culpa, por eso, a esta responsabilidad se le llama objetiva en ausencia del elemento subjetivo culpa.”⁷⁶

Para Gutiérrez y González, el maquinismo y los múltiples casos de afectaciones que produce su uso, aceleran a algunos autores para discutir el fundamento de este tipo de responsabilidad y toman la noción de riesgo en lugar del concepto culpa para darle soporte.⁷⁷ A decir de Gutiérrez y González, el penalista italiano Enrico Ferri (1856-1929)⁷⁸ llegó a sostener que él era el creador de la teoría de la responsabilidad sin culpa o responsabilidad objetiva por riesgo creado y comenta lo siguiente:

⁷⁶ Cfr. “RESPONSABILIDAD CIVIL POR HECHOS PROPIOS, AQUILIANA Y OBJETIVA. DIFERENCIAS.”. [TA]; 9a. Época; T.C.C.; S.J.F. y su Gaceta; Tomo IV, Noviembre de 1996; Pág. 512 II.10.C.T.85 C.

⁷⁷ Ernesto Gutiérrez y González, op. cit., p. 796.

⁷⁸ Enrico Ferri también sostenía que el libre albedrío era una ilusión subjetiva. En su tesis doctoral “La teoría de la imputabilidad y la negación del libre albedrío” (1878) sostuvo que el derecho penal no podía fundarse sobre las bases del libre albedrío, criticando a la Escuela Clásica del Derecho Penal y sugiriendo la “teoría de la peligrosidad” y la actividad psicofísica de la persona para la sanción penal. “Si el reato es ajeno a la libertad del querer, no hay ninguna razón para que el Estado, y por su conducto, la sociedad, permanezcan inactivos frente a acciones delictuosas en las que para los clásicos falta la libertad de querer. Nace entonces el criterio de la defensa social como fundamento del derecho penal. El hombre dice FERRI, es responsable siempre de todo acto que realice, solo porque y en cuanto vive en sociedad.” Cfr. Jesús Bernal Pinzón, “Imputabilidad, culpabilidad y responsabilidad”, *Nuevo Foro Penal*, no. 56, abril 1992, p. 262. <https://publicaciones.eafit.edu.co/index.php/nuevo-foro-penal/article/download/4033/3287>

“Hoy se esparce en Francia una teoría objetiva de la responsabilidad civil que, apoyándose en la misma idea introducida por mí en la teoría de la responsabilidad penal, o sea la de que esta responsabilidad es independiente de la culpa, confirma la razón común de la responsabilidad civil y penal. En el dominio jurídico-criminal, como en el dominio jurídico civil cualquier hombre, siempre y en cada caso, determina mediante cada una de sus reacciones una reacción social correspondiente: siempre, por consiguiente, experimenta las consecuencias naturales y sociales de sus propios actos, de los cuales es responsable, por el solo hecho de haberlos ejecutado”⁷⁹

Así fue -dice Gutiérrez y González- como empezaron a cobrar auge transportadas al terreno de la responsabilidad civil, las teorías del positivismo penal que Ferri, consideraba como novedosas, equivocadamente, y que no lograron éxito en el campo del Derecho Penal.⁸⁰

Existen casos en los que no habrá responsabilidad objetiva por riesgo creado, aunque se produzca un daño: 1. Si no hay relación de “causalidad” entre el daño y el objeto peligroso (el daño siempre debe ser consecuencia inmediata y directa del empleo del objeto peligroso, y *no imputable a tercera persona* que produzca una causa que lo lleva a provocar a su vez el daño⁸¹); 2. Si hay culpa de la víctima (no

⁷⁹ Cfr. Henri Mazeaud et. al. *Tratado Teórico y Práctico de la Responsabilidad Civil Delictual y Contractual. Compendio. Tomo I*, Buenos Aires, EJE, 1977, p. 27.

⁸⁰ Ernesto Gutiérrez y González, op. cit., p. 797. También la *Ley Federal del Trabajo* lo contempla en su artículo 465-bis: “El patrón es responsable de la seguridad e higiene y de la prevención de los riesgos en el trabajo, conforme a las disposiciones de esta Ley, sus reglamentos y las normas oficiales mexicanas aplicables. Es obligación de los trabajadores observar las medidas preventivas de higiene establecen los reglamentos y las normas oficiales mexicanas expedidas por las autoridades competentes, así como las que indiquen los patrones para la prevención de riesgos de trabajo.”. En principio, siguiendo nuevamente a Gutiérrez y González, podemos encontrar un dato histórico interesante: México inició la doctrina del riesgo creado desde la redacción del Código Civil de 1870. El artículo 1595 decía: “También habrá lugar a la responsabilidad civil por los daños que causen los establecimientos industriales, ya en razón del peso y movimiento de las máquinas...”. Posteriormente, encontramos el concepto de “riesgo creado” en la Constitución de 1917, al momento de la creación del artículo 123, inciso A, fracción XIV (del texto original): “Los empresarios serán responsables de los accidentes del trabajo y de las enfermedades profesionales de los trabajadores sufridas con motivo o en ejercicio de la profesión o trabajo que ejecuten...”⁸⁰ Y finalmente, el Código Civil de 1928, en su artículo 1913 comenta: “Artículo 1913.-Cuando una persona hace uso de mecanismos, instrumentos, aparatos o substancias peligrosas por sí mismos, por la velocidad que desarrollen, por su naturaleza explosiva o inflamable, por la energía de la corriente eléctrica que conduzcan o por otras causas análogas, está obligada a responder del daño que cause, aunque no obre ilícitamente, a no ser que demuestre que ese daño se produjo por culpa o negligencia inexcusable de la víctima.”.

⁸¹ Ernesto Gutiérrez y González, op. cit., p. 804.

será responsable, si se comprueba que el daño se provocó por culpa inexcusable de la víctima⁸²); 3. En caso fortuito; y, 4. Por prescripción.

Respecto a la categorización de la responsabilidad civil en contractual y extracontractual que la doctrina ha elaborado expondremos a continuación algunas diferencias importantes entre cada tipo de responsabilidad que sirven de guía para la toma de decisiones en las empresas. Juan Luis González Alcántara hace un ejercicio relevante de síntesis:

1. La capacidad para ser responsable civilmente, en materia contractual, requiere de capacidad general, en cambio en la otra un incapaz puede obligarse (artículo 1911, Código Civil).
2. Las cláusulas de limitación-exoneración de responsabilidad, solo son pensables en los contratos (artículo 2117, Código Civil).
3. Para conocer la responsabilidad contractual es juez competente el del lugar de cumplimiento, y a su falta, el del domicilio del demandado o el del lugar de la celebración del contrato; en cambio, el juez competente para conocer de la responsabilidad extracontractual (hecho ilícito) es el del lugar donde se produjo el daño. En nuestro Derecho es competente el juez del domicilio del demandado.
4. En la responsabilidad contractual cuando más de una persona incumple con la obligación asumida, si no hay pacto en contrario, responden en forma mancomunada, en tanto que cuando más de una persona comete un hecho ilícito civil, son responsables de modo solidario (artículo 1917, Código Civil).
5. La carga de la prueba corresponde al demandado al presumirse su culpa; en cambio en la responsabilidad extracontractual la prueba corresponde a la víctima demandante.
6. La extensión de la reparación del daño en la responsabilidad contractual solo comprende los previsibles, y en la extracontractual se extiende, además, a las consecuencias causales del hecho. En el Derecho mexicano sólo se responde de los daños que sean consecuencia directa e inmediata de la conducta ilícita.

⁸² *Idem.*

7. La mora sólo se presenta en la responsabilidad contractual; y,
8. La prescripción de las acciones derivadas de relaciones contractuales fenece a los diez años, salvo que la ley establezca un plazo especial, la extracontractual prescribe en dos años (artículo 1934, Código Civil)⁸³.

En el tema contractual, son varios los derechos que con motivo de la violación de un contrato asisten a la víctima y es de gran importancia conocerlos para la empresa y su área legal. Una parte de estos derechos los encontramos en el artículo 1949: *“La facultad de resolver las obligaciones se entiende implícita en las recíprocas, para el caso de que uno de los obligados no cumpliere lo que le incumbe. El perjudicado podrá escoger entre exigir el cumplimiento o la resolución de la obligación, con el resarcimiento de daños y perjuicios en ambos casos. También podrá pedir la resolución aún después de haber optado por el cumplimiento, cuando éste resultare imposible.”*

Otro derecho de una víctima por un hecho ilícito que parte del cumplimiento del clausulado de un contrato es la declaración de rescisión, es decir, aquel “acto unilateral, por el cual se le pone fin, salvo que la ley lo prohíba, de pleno derecho *ipso jure* -sin necesidad de declaración judicial- a otro acto bilateral, plenamente válido, por incumplimiento culpable, en este imputable a una de las partes.”⁸⁴. La rescisión forma parte de la resolución, acto jurídico por el cual: 1.- se priva de sus efectos, total o parcialmente para el futuro, a un acto jurídico anterior, plenamente válido, y, 2.- Los efectos pasados de este, siendo lícitos, pueden o no quedar subsistentes, según la naturaleza del acto, o la voluntad de las partes.⁸⁵

Los contratantes pueden poner cualquier tipo de cláusulas en un contrato. Por ello, los contratos deben ser medidos a través de la parametrización de su clausulado dentro de una matriz de riesgos. Las más comunes y que son reguladas por los códigos civiles tienen que ver con la estipulación de cierta prestación como

⁸³ Cfr. Juan Luis González Alcántara, *La responsabilidad civil de los médicos*, México, Porrúa, 2009, pp. 70 y 71.

⁸⁴ *Idem.*, pp. 642-643.

⁸⁵ Existe también la revocación o terminación (acto jurídico unilateral o bilateral, por medio del cual se pone fin a otro acto jurídico anterior, unilateral o bilateral, plenamente válido, por razones de conveniencia y oportunidad catalogadas subjetivamente por una sola parte, o bien apreciadas en forma objetiva por ambas según sea el caso) y la modificación común (unilateral o bilateral). Cfr. Gutiérrez y González, *op. cit.*, p. 643 y ss.

pena para el caso de que la obligación no se cumpla⁸⁶ y también está el tema del daño y perjuicio, contenidas en los artículos 2108, 2109 y 2110:

Artículo 2108.- Se entiende por daño la pérdida o menoscabo sufrido en el patrimonio por la falta de cumplimiento de una obligación.

Artículo 2109.- Se reputa perjuicio la privación de cualquiera ganancia lícita, que debiera haberse obtenido con el cumplimiento de la obligación.

Artículo 2110.- Los daños y perjuicios deben ser consecuencia inmediata y directa de la falta de cumplimiento de la obligación, ya sea que se hayan causado o que necesariamente deban causarse.⁸⁷.

Por lo general, los códigos civiles señalan que quien con su actuar u omisión de actuar cause un daño a otro, será responsable de daños y perjuicios, lo cual conlleva de manera natural a una indemnización por daños y perjuicios derivados de un hecho ilícito o por riesgo creado. Tradicionalmente, este tipo de responsabilidad trae aparejada una indemnización.

La siguiente tabla nos muestra cómo, en todas las categorías de responsabilidad civil, la legislación contempla la participación de una persona y su voluntad (participación directa o de una tercera persona), utilizando a la culpa, dolo, negligencia o el riesgo creado, como elementos para materializar el daño o perjuicio (indemnización):

⁸⁶ Ver artículo 1840 del Código Civil Federal mexicano.

⁸⁷ Siguiendo a Gutiérrez y González, por daño debe entenderse “la pérdida o menoscabo sufrido en el patrimonio de una persona, por conductas lícitas o ilícitas de otra persona, provocadas directamente por sí, o por no cuidar bien a personas bajo su custodia, o por cosas que posee ésta y que la ley considera para responsabilizar a su autor; y que por perjuicio se entiende la privación de cualquier ganancia lícita que debiera de haberse obtenido, de no haberse generado la conducta lícita o ilícita de otra persona, a que antes se hace referencia y que la ley considera para responsabilizarla.” Ernesto Gutiérrez y González, *op. cit.*, p. 744.

Fuente y destino del daño y excluyentes de responsabilidad

Origen del daño	Tipo de acto	Destino del daño	Excluyentes de responsabilidad
Persona	Hecho ilícito: Obra ilícitamente / contra buenas costumbres (Art. 1830)	Persona	Que se demuestre culpa o negligencia inexcusable de la víctima
Persona	Acto de incapaces, de sus tutores, del que lleva la patria potestad o de quien los cuida (directores de colegios, talleres, etc). (Art. 1920)	Persona	Que se demuestre que hicieron todo lo posible por evitar el daño a la víctima
Persona	Ejercicio de un derecho (Arts. 1919-1922)	Persona	Que ejerza el derecho sin intención de hacer daño a otro
Persona	Uso de máquinas, sustancias peligrosas, por la velocidad que desarrollen, por la energía de la corriente que conduzcan o cualquier otra causa análoga (Art. 1913)	Persona	Que se demuestre culpa o negligencia inexcusable de la víctima
Persona	Sin empleo de máquinas y demás supuestos y sin culpa o negligencia de ninguna de las partes. (Art. 1914)	Persona	Cada una de las partes soporta los daños sin derecho a indemnización
Persona	Afectación a sentimientos, afectos, creencias, decoro, honor, reputación, vida privada, etc. (Art. 1916)	Persona	La no acreditación de la ilicitud de la conducta y el daño que directamente hubiere causado tal conducta
Persona jurídica	Responsables por el actuar de sus representantes legales (Art. 1918)	Persona	Que se demuestre que hicieron todo lo posible por evitar el daño a la víctima
Persona jurídica	Actos realizados por sus trabajadores. (Art. 1924)	Persona	Que se demuestre que hicieron todo lo posible por evitar el daño a la víctima
Animal	Dueño responsable del daño de animales (Art. 1929)	Persona	Que lo guardaba y vigilaba; que fue provocado; que hubo imprudencia del ofendido

Fuente: El autor

Categorías de responsabilidad penal. - Hemos visto hasta aquí un panorama descriptivo de las categorías de responsabilidad civil. Ahora daremos un panorama descriptivo de las categorías de responsabilidad penal⁸⁸. Las categorías de responsabilidad en materia penal tienen relación con las concepciones subjetivas y objetivas que incluyen el dolo y la culpa, tiene que ver también con los tipos penales y la aplicación de la teoría del delito⁸⁹ que utilizan los practicantes del Derecho de cada país. La responsabilidad subjetiva, implica negligencia, dolo o intencionalidad en la realización del daño. La responsabilidad objetiva, tiene un origen civil y se apoya en la teoría del riesgo, donde hay ausencia de intencionalidad dolosa.⁹⁰ El dolo y la culpa son los elementos que marcan la categoría de responsabilidad penal.⁹¹

Por otro lado, la dogmática penal ha creado la llamada Teoría del Delito, que busca generar un camino lógico para fijar la responsabilidad penal (incriminación o imputación penal) y se constituye por la búsqueda de una conducta típica, antijurídica y culpable. El Derecho Penal, como el Derecho Civil toma en cuenta el libre albedrío como fundamento de la imputabilidad y como parte del juicio de culpabilidad que justifica la imposición de una pena. La Teoría del Delito es un instrumento para determinar la aplicación o no de la ley penal. La ley penal debe ser vista como un sistema que contiene premisas con definiciones, hipótesis y consecuencias (delito y sus elementos, tentativa, participación, autoría) que sirven para determinar la responsabilidad de una persona. Las siguientes dos tablas

⁸⁸ El principio de responsabilidad corporativa de las sociedades -dice Luigi Ferrajoli, y en general de las personas jurídicas es de origen germánico y ha dominado, casi sin discusión, durante la Edad Media hasta Bartolo. Aún en el siglo pasado y en el actual ha encontrado algún defensor como Gierke, von Liszt, Hafter, Longhi, De Marsico. Cfr. Luigi Ferrajoli, *Derecho y razón. Teoría del Garantismo penal*, Madrid, Trotta, 2009, p. 523.

⁸⁹ La teoría del delito, en el mundo latino, ha sido influenciada por la Filosofía alemana. Hay una teoría del delito basada en el causalismo o sistema clásico, en donde encontramos a pensadores como Carl Binding (Las normas y su transgresión, 1872), Ernst von Beling (La prohibición de las pruebas como límite de la verdad en la instrucción criminal, 1903), Von Liszt (La idea del fin en el derecho penal, 1882) y Edmund Mezger (Derecho penal, 1952). La teoría del finalismo, con Hans Welzel (Estudios sobre el sistema del derecho penal, 1939). Y, el funcionalismo de Claus Roxin (Teoría del Tipo Penal, tipos abiertos y elementos del deber jurídico (1979), Autoría y Dominio del Hecho en Derecho Penal (1998) y Günther Jakobs (Derecho penal. Parte general: fundamentos y teoría de la imputación (1997).

⁹⁰ Cfr. Luigi Ferrajoli, *Derecho y razón. Teoría del Garantismo penal*, Trotta, Madrid, 2009, pp. 488-489

⁹¹ Consúltense también las páginas 382 a 392 de este trabajo (Hart y Dworkin).

concentran los elementos del delito y una segmentación de elementos positivos y negativos.

Elementos del delito

Aspectos positivos	Aspectos negativos (impiden su integración)
Conducta (acción/omisión)	Ausencia de conducta
Tipicidad	Atipicidad
Antijuridicidad	Causas de justificación
Culpabilidad	Inculpabilidad
Punibilidad	Excusas absolutorias

Fuente: El autor

Integración de los elementos positivos//negativos del delito según el CPDF

Conducta típica (elementos del tipo o cuerpo de delito)	Antijuridicidad	Culpabilidad
Atipicidad - Por ausencia de conducta - Por falta de elementos del tipo penal - Por error de tipo - Por consentimiento disponibilidad del BJ	Causas de justificación - Legítima defensa - Estado de necesidad - Cumplimiento de un deber - Ejercicio de un derecho - Consentimiento presunto	Inculpabilidad - Estado de necesidad disculpante o exculpante - Inimputabilidad y acción libre en su causa - Error de prohibición - Inexigibilidad de otra conducta

Fuente: El autor.

La división de los elementos del delito tiene una finalidad de sistematización de los mismos para asegurar al operador del Derecho que realizará un análisis ordenado y completo de los hechos penalmente relevantes. Esto se hace de la siguiente manera:

- I. Analizar si existe una acción humana contenida en la norma;
- II. Analizar si la acción llena los elementos del tipo penal (si es típica);
- III. Analizar si la realización de la acción resulta justificada es antijurídica o la propia ley [de manera excepcional] contiene una causa de justificación;
- IV. Si se confirma que es una acción típica y antijurídica, entonces se analiza el grado de responsabilidad del autor (culpabilidad).

No se puede saltar un paso y pasar a otro. Todo debe ocurrir y estar concatenado para llegar a la culpabilidad.

Los responsables del delito según las formas de autoría y participación se clasifican en:

I.- Los que acuerden o preparen su realización.

II.- Los que los realicen por sí;

III.- Los que lo realicen conjuntamente;

IV.- Los que lo lleven a cabo sirviéndose de otro;

V.- Los que determinen dolosamente a otro a cometerlo;

VI. - Los que dolosamente presten ayuda o auxilien a otro para su comisión.

VII.- Los que con posterioridad a su ejecución auxilien al delincuente, en cumplimiento de una promesa anterior al delito; y

VIII.- Los que, sin acuerdo previo, intervengan con otros en su comisión, cuando no se pueda precisar el resultado que cada quien produjo.⁹²

Luigi Ferrajoli nos explica el concepto responsabilidad y el concepto culpabilidad. En cuanto al primero, nos aclara que por tal “ha de entenderse simplemente la sujeción jurídica a la sanción como consecuencia de un delito, cualquiera que sea la relación normativamente exigida entre el sujeto llamado a responder y el delito de que responde”⁹³.

Por su parte, el concepto de culpabilidad puede descomponerse en tres elementos, que “constituyen otras tantas condiciones subjetivas de responsabilidad en el modelo penal garantista: a) la personalidad de la acción, que designa la susceptibilidad de adscripción material del delito a la persona de su autor, esto es, la relación de causalidad que vincula recíprocamente decisión del reo, acción y resultado del delito; b) la imputabilidad o capacidad penal, que designa una condición psico-física del reo, consistente en su capacidad, en abstracto, de entender y de querer; c) la intencionalidad o culpabilidad en sentido estricto, que designa la consciencia y voluntad del concreto delito y que, a su vez, puede asumir la forma de dolo o de culpa, según la intención vaya referida a la acción y al resultado o sólo a la acción y no al resultado, no querido ni previsto sí previsible”⁹⁴.

⁹² Artículo 13 del Código Penal Federal mexicano.

⁹³ Cfr. Luigi Ferrajoli, *Derecho y razón. Teoría del Garantismo penal*, Madrid, Trotta, 2009, pp. 490.

⁹⁴ *Idem*.

Desde un enfoque epistémico -dice Ferrajoli- existen principios que componen el modelo del Derecho Penal garantista⁹⁵ que deben seguir las democracias constitucionales⁹⁶. Se trata de un modelo límite, solo tendencial y nunca perfectamente satisfacible. Su axiomatización resulta de la adopción de 10 axiomas o principios fundamentales:

Teoría del garantismo penal y sus diez principios

Nulla poena sine crimine	Retributividad
Nullum crimen sine lege	Legalidad
Nulla lex sine necessitate	Necesidad
Nulla necessitas sine iniuria	Lesividad
Nulla iniuria sine actione	Materialidad
Nulla actio sine culpa	Culpabilidad
	Jurisdiccionalidad
	Acusatorio
	Carga de la prueba
	Contradictorio

Las empresas deben medir sus riesgos penales con base en la legislación vigente del lugar en el que tengan su establecimiento y tomar en cuenta los tipos penales que tienen relación con su objeto social, con los productos y servicios que ofrecen, con su personal, socios, accionistas y toda aquella persona que tiene un vínculo. Las empresas deben tener el debido control en su organización para evitar la responsabilidad penal. El artículo 421 del Código Nacional de Procedimientos Penales señala:

⁹⁵ *Ibidem.*, pp. 93-116

⁹⁶ La democracia constitucional para Ferrajoli es “un sistema político complejo, basado en límites y vínculos jurídicos impuestos a todos los poderes -públicos y privados, estatales y supraestatales- para la tutela de la paz y de los derechos fundamentales de todos.” Cfr. *Principia iuris. Teoría del derecho y de la democracia, Tomo 1, Teoría del derecho*, Madrid, Trotta, 2011, contraportada.

Artículo 421. Ejercicio de la acción penal y responsabilidad penal autónoma.

Las personas jurídicas serán penalmente responsables, de los delitos cometidos a su nombre, por su cuenta, en su beneficio o a través de los medios que ellas proporcionen, cuando se haya determinado que además existió ***inobservancia del debido control en su organización***. Lo anterior con independencia de la responsabilidad penal en que puedan incurrir sus representantes o administradores de hecho o de derecho.

El Ministerio Público podrá ejercer la acción penal en contra de las personas jurídicas con excepción de las instituciones estatales, independientemente de la acción penal que pudiera ejercer contra las personas físicas involucradas en el delito cometido.

No se extinguirá la responsabilidad penal de las personas jurídicas cuando se transformen, fusionen, absorban o escindan. En estos casos, el traslado de la pena podrá graduarse atendiendo a la relación que se guarde con la persona jurídica originariamente responsable del delito.

La responsabilidad penal de la persona jurídica tampoco se extinguirá mediante su disolución aparente, cuando continúe su actividad económica y se mantenga la identidad sustancial de sus clientes, proveedores, empleados, o de la parte más relevante de todos ellos.

Las causas de exclusión del delito o de extinción de la acción penal, que pudieran concurrir en alguna de las personas físicas involucradas, no afectará el procedimiento contra las personas jurídicas, salvo en los casos en que la persona física y la persona jurídica hayan cometido o participado en los mismos hechos y estos no hayan sido considerados como aquellos que la ley señala como delito, por una resolución judicial previa. Tampoco podrá afectar el procedimiento el hecho de que alguna persona física involucrada se sustraiga de la acción de la justicia.

Las personas jurídicas serán penalmente responsables únicamente por la comisión de los delitos previstos en el catálogo dispuesto en la legislación penal de la federación y de las entidades federativas.

En este tema queremos destacar las aportaciones de la Teoría de la Imputación Objetiva y que ha intentado revolucionar la teoría del delito en muchas partes del mundo occidental y ha influido en el modo de entender la tipicidad y la responsabilidad del médico al momento de realizar el acto médico y ha sido seguida por diferentes tribunales en el mundo, para resolver casos médicos cuya solución está en la norma penal y en la posibilidad de atribuir o no resultados basados en lo causal (la existencia de riesgos, su materialización y los resultados lesivos; la

causalidad como concepto ontológico físico-natural). Roxin⁹⁷, Welzel⁹⁸, English⁹⁹, Gimbernat¹⁰⁰, Jakobs¹⁰¹ y Frisch¹⁰² son los autores que han aportado -directa o indirectamente- elementos para delimitar la teoría de la imputación objetiva y su uso para determinados tipos penales. Enrique Gimbernat Ordeig nos comenta:

“Vuelvo a cuál es el contenido y la razón de ser de la teoría de la imputación objetiva. Esta sólo entra en juego cuando se ha comprobado que el autor doloso o imprudente ha causado (condicionado) un resultado típico. Porque si la conducta no es dolosa ni imprudente o porque, siéndolo, no ha causado el resultado típico, entonces ello basta y sobra ya para afirmar la no-tipicidad de ese comportamiento, puesto que en él falta alguno de los elementos subjetivos (dolo o, en su caso, imprudencia) u objetivos (relación de causalidad) de la tipicidad. Sólo cuando se ha constatado la existencia de un comportamiento doloso o imprudente causante (condicionante) de un resultado típico es cuando se plantea (como último requisito de la tipicidad objetiva) si ese resultado típico le es imputable objetivamente, o no, a quien lo ha causado con dolo o imprudencia. (En el caso del comportamiento doloso, si el resultado no se ha causado, entonces es obvio que no se puede imputar objetivamente un resultado no sobrevenido, sino que únicamente podría entrar en juego la eventual responsabilidad del autor por una tentativa de delito) ... Las primeras críticas que se formularon contra la teoría de la imputación objetiva procedían de la escuela finalista. Pero ello obedeció, fundamentalmente, a que la teoría de la imputación objetiva iba desplazando, poco a poco, como tal vez el tema central de la teoría del delito, a la discusión sobre el concepto de acción, algo que, comprensiblemente, esa escuela se resistía a aceptar. En cualquier caso, lo cierto es que –para disgusto de la escuela finalista– los miles de páginas que antes se dedicaban a la teoría de la acción han sido desplazadas, actualmente, por los igualmente miles de páginas que hoy tienen como materia de estudio a la imputación objetiva. Algo que, como he tratado de demostrar en algunos de los artículos que se recogen en el presente volumen, es *la ubicación correcta del problema*, pues ni el concepto predominantemente ontológico de la acción, ni el exclusivamente ontológico de la causalidad, pueden resolver satisfactoriamente casos como los que integran la imputación objetiva, que sólo pueden encontrar una respuesta –y

⁹⁷ Claus, Roxin, *Derecho Penal. Parte General. Fundamentos. La estructura de la teoría del delito*, Madrid, Editorial Civitas, 1997.

⁹⁸ Hans Welzel, *El nuevo sistema del derecho penal. Una introducción a la teoría de la acción finalista*, Buenos Aires, Editorial B de F, 2004; y también *Derecho Penal Alemán. Parte General*, Santiago de Chile, Editorial Jurídica de Chile, 1970.

⁹⁹ Karl Engisch, *La teoría de la libertad de la voluntad en la actual doctrina filosófica del derecho penal*, Buenos Aires, Editorial B de F, 2006.

¹⁰⁰ Enrique Gimbernat Ordeig ha escrito mucho sobre este tema. En particular se recomienda: 1) *Casualidad, omisión e imprudencia*, Anuario de derecho penal y ciencias penales, Tomo 47, Fasc/Mes 3, Madrid, 1994, págs. 5-60; 2) *En defensa de teoría de la imputación objetiva contra sus detractores y, también, contra algunos de sus partidarios*, Anuario de derecho penal y ciencias penales, Tomo 73, Fasc/Mes 1, 2020, Madrid, pp. 9-20.

¹⁰¹ Günther Jakobs, *La imputación objetiva en el derecho penal*, Buenos Aires, Editorial Ad-Hoc, 1997.

¹⁰² Wolfgang Frisch, *Comportamiento típico e imputación del resultado*, Madrid, Marcial Pons, 2004; y también: *Tipo penal e imputación objetiva*, Editorial B de f., Montevideo, 2020.

fundamentación– adecuadas sobre la base de los *criterios normativos* con los que opera dicha teoría.”¹⁰³

En la teoría de la imputación objetiva hay que distinguir entre aquellas causas que generan los individuos y aquellas causas que ocurrieron, pero que no se les puede atribuir a los individuos; “... ontológicamente, a nivel de lo que sucede en la naturaleza, todas estas situaciones serían realmente causas, pero lo procedente es determinar cuál de estas causas es la que tiene la relevancia jurídica, a cuál de ellas le corresponde el resultado, a cuál de ellas le es imputable el resultado”¹⁰⁴

Roxin y Jakobs son los autores que más han impulsado el uso de la teoría de la imputación objetiva. El primero, se apoya en sumar a la teoría del delito (dogmática) y la llamada política criminal, así como utiliza la teoría del ámbito de protección de la norma, la teoría de la prohibición de regreso, así como el supuesto de excluir el resultado lesivo para el bien jurídico, cuando la víctima es la responsable de atender determinados consejos -como los cuidados médicos- sirven para excluir de la imputación objetiva¹⁰⁵.

Los conceptos de la imputación objetiva se toman de la norma jurídica y su misión reguladora de la conducta humana (mandatos – prohibiciones). Los mandatos exigen acciones y las prohibiciones evitan resultados no deseados. En los dos supuestos, se trata de resultados “que sean posibles de ser producidos o evitados para la persona que se encuentra en una situación que el legislador presupone”¹⁰⁶, pero no habla de resultados generados por sistemas de inteligencia artificial en los que no participa ¹⁰⁷el ser humano. Aquí la teoría de la imputación objetiva se refiere a resultados que las personas pueden prever, evitar o producir.

Lo interesante de la imputación objetiva es la incorporación del concepto riesgos, reconociendo que todo en la vida del hombre -desde su propio nacimiento-

¹⁰³ Cfr. Enrique Gimbernat, “En defensa de teoría de la imputación objetiva contra sus detractores y, también, contra algunos de sus partidarios”, *Anuario de derecho penal y ciencias penales*, Tomo 73, Madrid, Fasc/Mes 1, 2020, pp. 9-20.

¹⁰⁴ Cfr. Francisco Bernate Ochoa, *Imputación objetiva y responsabilidad penal médica*, Bogotá, Universidad del Rosario, 2010, p. 47.

¹⁰⁵ Para conocer más detalles sobre estos conceptos consúltese a Francisco Bernate Ochoa, op. cit., pp. 58-89.

¹⁰⁶ *Ibidem.*, p. 60.

¹⁰⁷ *Ibidem.*, p. 61.

está rodeado de riesgos. Los riesgos no pueden desaparecer, solo se pueden mitigar o reducir mediante acciones preventivas. Debemos aprender a vivir con riesgos y regularlos, es imposible terminar con ellos.

En el caso de Jakobs y su propuesta de Teoría de la imputación objetiva, Bernate Ochoa, comenta que “los parámetros del riesgo permitido, el principio de confianza, las acciones a propio riesgo y el ámbito de la víctima son los que determinan los ámbitos de competencia, con lo que la imputación objetiva absorbe toda la teoría del delito, dejando de ser, como en la propuesta de Roxin un mecanismo corrector de la causalidad a nivel de tipo objetivo, para convertirse en el escenario donde se agota la teoría del delito”.¹⁰⁸

La norma da los criterios para estructurar los contenidos y límites de los riesgos permitidos y demás ingredientes que menciona Bernate Ochoa. Sin embargo, Jakobs tampoco logra encuadrar una conducta en la que no hay participación del ser humano en la materialización de los riesgos. La teoría de la imputación objetiva requiere forzosamente de la acción u omisión del ser humano (conducta-voluntad) y permite hacer una jerarquización de la responsabilidad de las personas que participaron directa o indirectamente en el delito (no todo es responsabilidad de todos, hay que calificar cada conducta). Este tipo de teoría ha dado buenos resultados en la persecución y sentencia de delitos relacionados con hechos de corrupción, en los que la responsabilidad de los sujetos involucrados se busca diluir entre tantas voluntades que se ven involucradas de los operadores o a los que materializan el acto y, por lo general, buscan que los líderes, cabezas del actuar corruptor no aparezcan vinculados con alguna responsabilidad (en jerarquía pudieran ser los superiores jerárquicos del servidor público que materializa el hecho de corrupción).

Aquí también se da el dilema conocido como *The many hands problem*, es decir, cuando muchas personas participan en una actividad, es difícil, sino imposible, determinar quién es el responsable y qué hacer cuando la responsabilidad no es rastreable por la propia naturaleza del sistema y por las

¹⁰⁸ Cfr. Francisco Bernate Ochoa, *op. cit.*, p. 70.

propias categorías jurídicas existentes.¹⁰⁹ Aquí las categorías de responsabilidad se estresan y entran en crisis. Entremos ahora a detalle con el tema de la responsabilidad del médico y su relación con el Derecho Civil, penal y administrativo.

La mayoría de los sistemas informáticos que se utilizan hoy en día son productos no de programadores individuales que trabajan de forma aislada, sino de grupos u organizaciones, normalmente corporaciones. Estos grupos, que con frecuencia reúnen equipos de personas con una amplia gama de habilidades y variados grados de experiencia, pueden incluir diseñadores, ingenieros, programadores, escritores, psicólogos, artistas gráficos, gerentes y vendedores. En consecuencia, cuando un sistema funciona mal y da lugar a daños, la tarea de asignar responsabilidades, el problema de identificar quién es responsable, se exagera y se oscurece. La responsabilidad característicamente entendida y analizada tradicionalmente en términos de un solo individuo, no se generaliza fácilmente a la acción colectiva. En otras palabras, mientras que la búsqueda más simple de la rendición de cuentas nos dirigiría en busca de aquel que debe dar un paso adelante (por ejemplo, el chico de la última fila contesta por la bola de escupir), la acción colectiva presenta un desafío. El análisis de la culpa, en términos de causa y culpa, puede ayudar a aclarar cómo, en los casos de acción colectiva, la rendición de cuentas se puede perder, o al menos, oscurecer.”¹¹⁰

¹⁰⁹ Ibo van de Poel et. al., comenta que quizás fue Denise Thompson, el primero en utilizar la noción “*The many hands problem*” en un artículo sobre la responsabilidad de los funcionarios públicos, lo describe de la siguiente manera: “Debido a que muchos funcionarios diferentes contribuyen de muchas maneras a las decisiones y políticas del gobierno, es difícil, incluso en principio, identificar quién es moralmente responsable de los resultados políticos” Cfr. van de Poel, I., Fahlquist, J. N., Doorn, N., Zwart, S., & Royakkers, L. (2012). The problem of many hands: climate change as an example. *Science and engineering ethics*, 18(1), 49–67, p. 50. Consúltese en: <https://doi.org/10.1007/s11948-011-9276-0> [20 de julio de 2022] Para profundizar más sobre este tema puede consultarse: a Ibo van de Poel, Lambèr Royakkers, Sjoerd D. Zwart, *Moral Responsibility and the Problem of Many Hands*, Routledge, 2018.

¹¹⁰ Cfr. Nissenbaum, H, Accountability in a computerized society en *Science and Engineering Ethics*, 2(1),25–42, 1996, p. 28, en este ensayo se argumenta sobre la erosión de la responsabilidad en las sociedades informatizadas. Argumenta que las computadoras crean barreras a la responsabilidad, basándose en análisis filosóficos de la culpa moral y la responsabilidad, se identifican cuatro barreras: (1) el problema de muchas manos, (2) el problema de los errores, (3) el problema de culpar a la computadora y (4) el problema de la propiedad del software sin responsabilidad. La autora dice “Most computer systems in use today are the products not of single programmers working in isolation but of groups or organizations, typically corporations. These groups, which frequently bring together teams of individuals with a diverse range of skills and vary in degrees of expertise, might include

Categorías de responsabilidad específicas de los médicos.- Algunas de las categorías de responsabilidad civil y penal vistas en las secciones anteriores, se van a utilizar para determinar la responsabilidad del médico. Este tema debe ser visto desde la Constitución Política de los Estados Unidos Mexicanos, la Ley General de Salud y la legislación civil, penal y administrativa que regula el actuar del médico o el acto médico. La razón primordial por la que la conducta del médico se regula de manera específica, a diferencia de otras profesiones, es que tiene conexión con una obligación de todo Estado, que es la salud pública, los servicios de salubridad y el derecho humano a la salud. Estas actividades, así como la libertad en el ejercicio profesional del médico basada en la Ética y la *Lex artis ad hoc*.¹¹¹ hacen que este tema sea tratado con una alta especialidad.

El ministro de la Suprema Corte de Justicia de la Nación, Juan Luis González Alcántara, fue de los pioneros en reflexionar sobre el derecho a la salud y la responsabilidad civil de los médicos, coincidimos cuando afirma que: “1) El Derecho a la salud es indispensable para el ejercicio de otros derechos. 2) El derecho a la salud requiere de medidas adecuadas para su protección. 3) La protección del derecho a la salud es una obligación impostergable del estado, de atención prioritaria. 4) El derecho a la salud trasciende directamente en la calidad de vida. 5) El derecho a la salud es autónomo y su protección no se limita a la abstención del

designers, engineers, programmers, writers, psychologists, graphic artists, managers, and salespeople. Consequently, when a system malfunctions and gives rise to harm, the task of assigning responsibility – the problem of identifying who is accountable – is exacerbated and obscured. Responsibility characteristically understood and traditionally analyzed in terms of a single individual, does not easily generalize to collective action. In other words, while the simplest quest for accountability would direct us in search of “the one” who must step forward (for example, the boy in the back row answering for the spit-ball), collective action presents a challenge. The analysis of blame, in terms of cause and fault, can help to clarify how in cases of collective action accountability can be lost, or at least, obscured.”

¹¹¹ El acto médico se debe efectuar con conocimiento de la especialidad sobre la que se decide y utilizando las habilidades y destrezas que la práctica profesional y la literatura médica indican (y acepta). *Lex Artis Ad Hoc* es el conjunto de reglas y conocimientos generados para el ejercicio de una especialidad médica, contenidos en distintos medios de almacenamiento, conservación y consulta acerca de técnicas y procedimientos que han sido universalmente aceptados y que se basan en los principios científicos y éticos que orientan la práctica médica. Cfr. Consejo Nacional de Salud Pública, A. C., *Lex Artis ad Hoc. Especialidades de Salud Pública, Epidemiología y Medicina Preventiva*, México, 2018. [lexartis.cdr \(cnsaludpublica.com.mx\)](http://lexartis.cdr.cnsaludpublica.com.mx) [20 de julio de 2022]

daño, sino a la exigencia de su protección. 6) El derecho a la salud está ligado al concepto de derechos esenciales.”¹¹²

La legislación mexicana, a partir de los artículos 4 y 5 de la Constitución, va mostrando conceptos claves para entender la responsabilidad médica, y a la vez, el derecho de toda persona a la protección de su salud.

El artículo 4 de la Constitución Política de los Estados Unidos Mexicanos señala: *“Toda persona tiene derecho a la protección de la salud. La Ley definirá las bases y modalidades para el acceso a los servicios de salud y establecerá la concurrencia de la Federación y las entidades federativas en materia de salubridad general, conforme a lo que dispone la fracción XVI del artículo 73 de esta Constitución. La Ley definirá un sistema de salud para el bienestar, con el fin de garantizar la extensión progresiva, cuantitativa y cualitativa de los servicios de salud para la atención integral y gratuita de las personas que no cuenten con seguridad social.”*

El artículo 5 de la misma Constitución dice: *“A ninguna persona podrá impedirse que se dedique a la profesión, industria, comercio o trabajo que le acomode, siendo lícitos. Nadie puede ser privado del producto de su trabajo, sino por resolución judicial. La ley determinará en cada entidad federativa, cuáles son las profesiones que necesitan título para su ejercicio, las condiciones que deban llenarse para obtenerlo y las autoridades que han de expedirlo... El contrato de trabajo sólo obligará a prestar el convenido por el tiempo que fije la ley... La falta de cumplimiento de dicho contrato, por lo que respecta al trabajador, sólo obligará a éste a la correspondiente responsabilidad civil...”*

En todo esto: ¿Cómo debemos entender el acto médico y cuáles son los elementos para calificar si tuvo o no responsabilidad?

El acto médico, materializa la relación médico - paciente. Es decir, es aquel “conjunto de acciones que van desde un simple diagnóstico hasta la más compleja intervención quirúrgica, y que lleva una sobre carga de riesgos, en la medida que

¹¹² Cfr. Juan Luis González Alcántara, *La responsabilidad civil de los médicos*, México, Porrúa, 2009, p. 3.

no se puede contar con elementos concretos y objetivos que permitan dar un resultado exacto”.¹¹³.

El acto médico requiere, por un lado, de la voluntad del paciente de acudir al médico y atender los síntomas que le afectan y, por el otro, de la voluntad del médico que busca atender esa petición. Para que el acto médico se materialice debe existir el conocimiento y la preparación necesaria, así como atender la llamada *Lex artis ad hoc*. El médico que lo realiza debe estar consciente que su intervención puede traer consecuencias en caso de no conocer la técnica adecuada o en caso de no contar con los conocimientos suficientes y actualizados para su desempeño profesional.

Ahora bien, si revisamos las categorías de responsabilidad, vistas desde el enfoque del constitucional y su relación con el ejercicio de alguna actividad profesional, ya sea privada o pública, individual o colectiva, en la que pudiera tener participación un médico, la Constitución es clara al mostrar cuatro tipos de responsabilidad en la que puede incurrir un médico:

1. Responsabilidad civil por el ejercicio profesional (artículo 5), en la que el médico por su cuenta o a través de algún hospital (público o privado) ejerce su profesión con base en los artículos 4 y 5 de la misma Constitución, pero por alguna circunstancia su actuar daña la salud de un paciente;

2. Responsabilidad civil derivada de la comisión de algún delito (artículo 22), en la que el médico por su cuenta o a través de algún hospital (público o privado) ejerce su profesión con base en los artículos 4 y 5 de la misma Constitución, pero por alguna circunstancia su actuar encuadra en algún tipo de conducta típica, antijurídica, culpable y punible por la ley penal;

3. Responsabilidad administrativa de aquellos médicos que son servidores públicos o aquellos particulares -de oficio médico- vinculados con alguna falta administrativa grave o hechos de corrupción (siguiendo las reglas y conceptos del Título Cuarto); y,

¹¹³ Enrique Orrego Puelles, “Mal praxis médica”, en *Acta cancerológica*, Vol. 31. No. 1, mayo 2002 citado por Sonia B. Fernández Cantón, Boletín CONAMED- OPS, marzo – abril 2016, México, p. 3. Puede consultarse en: http://www.conamed.gob.mx/gobmx/boletin/pdf/boletin5/acto_medico.pdf (20 de julio de 2022)

4. Responsabilidad patrimonial del Estado (109, último párrafo), como una vía procesal especial que tienen todas las personas que se sienten afectadas por el actuar del gobierno o por las omisiones del gobierno en los servicios de salud que ofrecen o a que tienen derecho y en el comportamiento del médico, sus actos y los daños o afectación producidos en su salud.¹¹⁴

¿Qué dice la Ley General de Salud en este último tema? En este último tema, el acceso a los servicios de salud que el gobierno proporciona deben alcanzar: 1) El bienestar físico y mental de las personas para el pleno ejercicio de sus capacidades; 2) La prolongación y mejoramiento de la calidad de vida; 3) La protección y el acrecentamiento de los valores que coadyuven a la creación, conservación y disfrute de condiciones de salud que contribuyan al desarrollo social; 4) El conocimiento para el adecuado aprovechamiento y utilización de los servicios de salud; 5) El desarrollo de la enseñanza y la investigación científica y tecnológica para la salud; y 6) La promoción de la salud y la prevención de las enfermedades.¹¹⁵

La salubridad general entra en uno de los tres tipos de servicios de salud que son regulados por el Estado y se define como el conjunto de servicios que se proporcionan al individuo, con el fin de proteger, promover y restaurar su salud. Los profesionales de salud se apoyan en las Guías de Práctica Clínica y los medios electrónicos de acuerdo con las normas oficiales mexicanas que al efecto emita la Secretaría de Salud.¹¹⁶

A nivel de la Ley General de Salud, el perfil profesional y académico del médico que realizó el acto y, en su caso, la afectación o daño debe analizarse para evaluar y emitir una resolución sobre la responsabilidad o no del sujeto ya sea en el diagnóstico o tratamiento de enfermedades (actividades médicas curativas). El artículo 33 de la *Ley General de Salud* estipula que para la realización de cualquier procedimiento médico quirúrgico de especialidad, los profesionales que lo ejerzan

¹¹⁴ Se entiende por salud aquel estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades. Cfr. *Ley General de Salud*, publicada en el Diario Oficial de la Federación (última reforma del 1 de junio de 2021) http://www.diputados.gob.mx/LeyesBiblio/pdf/142_010621.pdf

¹¹⁵ Cfr. *Ley General de Salud*, publicada en el *Diario Oficial de la Federación* (última reforma del 1 de junio de 2021) http://www.diputados.gob.mx/LeyesBiblio/pdf/142_010621.pdf

¹¹⁶ Artículo 24 y 32 de la *Ley General de Salud*.

requieren de: 1) Cédula de especialista legalmente expedida por las autoridades educativas competentes, es decir, no cualquier persona que haya estudiado medicina puede hacer un procedimiento quirúrgico solo aquellos que demuestren esta formalidad; y 2) Certificado vigente de especialista que acredite capacidad y experiencia en la práctica de los procedimientos y técnicas correspondientes en la materia, de acuerdo con la *Lex Artis Ad Hoc* de cada especialidad, expedido por el Consejo de la especialidad según corresponda.¹¹⁷

Esto tiene una razón de ser: La medicina no es una ciencia exacta. A diferencia de lo que se puede pensar, el diagnóstico médico se basa mucho en opiniones basadas en la heurística o en el razonamiento abductivo más que en método científico, por tanto, el resultado del diagnóstico y/o tratamiento puede salirse de control o de los parámetros de resultados que se esperaban. Para guiar el actuar médico, se han construido diversos postulados para su regulación, basados en Ética, Moral y Derecho, así como en el respeto a la persona (no dañar) y la búsqueda del bien (hacer el bien).

Estos elementos son tomados en cuenta por nuestro sistema jurídico construyendo lo que conocemos como la responsabilidad por el acto médico o la responsabilidad profesional del médico. También, estos elementos marcan la obligación del médico en su formación y en la necesidad de contar con reglas, procedimientos y prácticas que permitan llevar a cabo un proceso de evaluación del acto médico para responder una pregunta clave:

¹¹⁷ El artículo 81 de la *Ley General de Salud* regula estos temas: “Artículo 81.- La emisión de los diplomas de especialidades médicas corresponde a las instituciones de educación superior y de salud oficialmente reconocidas ante las autoridades correspondientes. Para la realización de los procedimientos médicos quirúrgicos de especialidad se requiere que el especialista haya sido entrenado para la realización de los mismos en instituciones de salud oficialmente reconocidas ante las autoridades correspondientes. El Comité Normativo Nacional de Consejos de Especialidades Médicas tendrá la naturaleza de organismo auxiliar de la Administración Pública Federal a efecto de supervisar el entrenamiento, habilidades, destrezas y calificación de la pericia que se requiere para la certificación y recertificación de la misma en las diferentes especialidades de la medicina reconocidas por el Comité y en las instituciones de salud oficialmente reconocidas ante las autoridades correspondientes. Los Consejos de Especialidades Médicas que tengan la declaratoria de idoneidad y que estén reconocidos por el Comité Normativo Nacional de Consejos de Especialidades Médicas, constituido por la Academia Nacional de Medicina de México, la Academia Mexicana de Cirugía y los Consejos de Especialidad miembros, están facultados para expedir certificados de su respectiva especialidad médica. Para la expedición de la cédula de médico especialista las autoridades educativas competentes solicitarán la opinión del Comité Normativo Nacional de Consejos de Especialidades Médicas.”.

Tanto el médico como la persona encargada de emitir un fallo respecto su responsabilidad jurídica, requiere de todos estos elementos para conocer cómo se hizo el diagnóstico, cuál fue procedimiento o tratamiento indicado, bajo qué criterio se decidió las dosis de medicamentos o la técnica quirúrgica utilizada para el caso específico; y, si existió supervisión durante todo el proceso -por el superior u otro experto- así como la documentación de todo lo ocurrido (expediente médico) y, en ocasiones, si se detectó algún daño no previsto que traiga como resultado una afectación adicional o no prevista en la salud de la persona.

Además de analizar la calidad del sujeto o el perfil de la formación del médico, existe otro elemento para tomar en cuenta y es denominado mala *praxis*. ¿Qué es una mala *praxis* médica?

En México, el Poder Judicial de la Federación ha tratado en reiteradas ocasiones el tema de la mala práctica y el error médico. En principio, debemos reiterar que la medicina no es una ciencia exacta basada en el método científico que permita diagnosticar y generar un tratamiento médico, por lo que es común en la práctica médica que existan diagnósticos divergentes o equivocados. Pero una cosa es una equivocación por error humano -*humanum errare est*- y otra cosa es una mala práctica. La jurisprudencia de los tribunales colegiados de circuito señala:

“...la medicina no es una ciencia exacta, pues existen matices en el curso de una enfermedad o padecimiento que pueden inducir a diagnósticos y terapias equivocadas. Así, una interpretación errada de los hechos clínicos por parte del médico puede llevar a un diagnóstico erróneo que, aunque no exime de responsabilidad al médico que lo comete, no reviste la gravedad de la negligencia médica, por lo que se habla entonces de un error excusable, pues lo que se le pide al “buen médico” es aplicar sus conocimientos, habilidades y destrezas con diligencia, compartiendo con el paciente la información pertinente, haciéndole presente sus dudas y certezas, tomando en cuenta sus valores y creencias en la toma de decisiones y actuando con responsabilidad y prudencia. Por otro lado, el término mal *praxis* (mala práctica médica) se ha acuñado para señalar conductas impropias del profesional frente a un paciente y que no sigue las normas o pautas que señala la *lex artis* médica, pero no hay aquí un error de juicio, sino que, o la actuación del médico que está en posesión de conocimientos y habilidades no ha sido diligente, o éste ha actuado con impericia e imprudencia frente a una situación clínica para la cual no está capacitado; este tipo de conducta médica constituye un error médico inexcusable y el profesional debe responder por esta conducta inapropiada. Por tanto, la responsabilidad profesional está subordinada a la previa acreditación de una clara negligencia en la prestación de los servicios, independientemente del resultado.”¹¹⁸

¹¹⁸ Cfr. “RESPONSABILIDAD PROFESIONAL MÉDICA. DISTINCIÓN ENTRE ERROR Y MALA PRÁCTICA PARA EFECTOS DE SU ACREDITACIÓN.”. [TA]; 10a. Época; T.C.C.; S.J.F. y su

La Primera Sala de la Suprema Corte de Justicia de la Nación complementa lo anterior diciendo:

En los casos de **responsabilidad civil de los hospitales privados**, derivada de la negligencia de sus médicos, resulta excesivo establecer que el usuario del servicio de salud debe demostrar la relación laboral o profesional entre aquéllos y el hospital para que proceda la responsabilidad civil de ambos, pues ***no sólo llevaría la carga de ser víctima de la mala praxis o del acto que motivó el daño, sino que judicialmente sería revictimizado, al obligársele a probar una cuestión fuera de su alcance***. En esas condiciones, los usuarios de los servicios de atención médica, así como sus familiares, están en una condición de desventaja por el desconocimiento del personal del hospital que tiene la calidad de empleado y la de independiente, pues no están enterados de las complejidades técnicas de los acuerdos contractuales y de empleo entre el hospital y el personal que opera ahí, al contrario del hospital, que sí tiene conocimiento y, además, decide cómo organizarse y representarse. De ahí que el usuario de los servicios de salud privada, al estar en una posición de desventaja frente a la institución médica, no tiene la carga de la prueba.¹¹⁹

En el acto médico y sus elementos está la clave de la responsabilidad del médico, al punto que se ha estudiado por la Suprema Corte de Justicia de la Nación, que existen actos médicos simples y actos médicos con una complejidad que amerita un estudio detallado de sus elementos: La Primera Sala de la Suprema Corte de Justicia de la Nación señala:

...el acto médico es una actuación compleja que debe ser examinada en su conjunto a fin de valorar la presencia o no de una conducta dolosa que actualice la responsabilidad civil, dado que los distintos actos de un profesionista o de varios profesionistas médicos se encuentran estrechamente vinculados. En ese tenor, ***cuando el acto médico está integrado por distintas facetas en las que participan sucesivamente una multiplicidad de profesionistas médico-sanitarios, se estima que puede atribuírsele una conducta negligente a un profesionista si de acuerdo con la lex artis ad hoc, se encontraba en una posición óptima para advertir el error culposos o doloso del profesionista o profesionistas que le precedieron y, ante tal situación, omitió llevar a cabo los actos necesarios para corregirlo o intentar corregirlo, siempre que tal actuar se encontrara dentro de su campo de acción en el acto médico***. Lo anterior no implica que un profesionista médico-sanitario sea siempre responsable por la actuación dolosa o negligente del que le precedió en el acto médico, pues ello podría

Gaceta; Libro XXV, octubre de 2013; Tomo 3; Pág. 1890 I.4o.A.64 A (10a.) Registro digital: 2004785. [El énfasis es propio]

¹¹⁹ “RESPONSABILIDAD CIVIL DE LOS HOSPITALES PRIVADOS DERIVADA DE LA NEGLIGENCIA DE SUS MÉDICOS. CARGA DE LA PRUEBA.” [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 16, marzo de 2015; Tomo II; Pág. 1112. [El énfasis es propio]
1a. CXVII/2015 (10a.) Registro digital: 2008747

actualizar una violación a los principios de legalidad y seguridad jurídica de esas personas. Más bien, **la atribución de responsabilidad a un profesionalista médico-sanitario que actúa en un acto médico complejo en el que intervinieron sucesivamente una serie de personas, dependerá enteramente de sus propios actos en dicha atención médica.** Así, por ejemplo, si tras la participación de un médico, otro médico convalida la negligencia del que le precedió u omite cumplir con sus deberes de cuidado que lo llevarían a intentar corregirlo o a informarlo a los responsables para hacerlo, según la *lex artis* médica, **tal conducta puede considerarse a su vez como culposa en sentido amplio y dar pie a una responsabilidad si se acredita que tal acto propició o incidió directamente en la producción del daño.**¹²⁰

En resumen, una mala práctica médica es una decisión del médico, basada en la autonomía de la voluntad y su razonamiento, que se toma fuera del deber de cuidado siguiendo el principio de la debida diligencia, ya sea exigida por el propio acto médico (especialidad), por la ley o por la *lex artis ad hoc*.

La realización de uno o más actos médicos con errores injustificados, trae como resultado la aplicación de la sanción que la norma jurídica contiene ya sea por negligencia, dolo o impericia. Las definiciones de estos supuestos las hemos puesto en la siguiente tabla:

Mala praxis médica (categorías)

Negligencia	Descuido de precauciones y atenciones calificadas como necesarias; cuando a pesar de tener el conocimiento no se aplica y se provoca daño.
Impericia	Falta de capacidad y experiencia, de conocimientos técnicos y prácticos; cuando se actúa sin tener el conocimiento y se genera daño.
Dolo	Maquinación o artificio para dañar a otro. Siempre es punible ya que viola conscientemente la ley.

Fuente: El autor con base en Sonia B. Fernández Cantón, Boletín CONAMED- OPS, marzo – abril 2016, México. Consúltase en: http://www.conamed.gob.mx/gobmx/boletin/pdf/boletin5/acto_medico.pdf

Para que haya mal praxis se debe comprobar la culpa del médico en su actuación. En cambio, el error médico puede ser por diversos factores: 1) de tipo institucional, como la burocracia o la carencia de recursos (humanos, materiales,

¹²⁰ “RESPONSABILIDAD CIVIL EXTRACONTRACTUAL EN MATERIA MÉDICO-SANITARIA. SUPUESTOS DE VALORACIÓN EN UN ACTO MÉDICO COMPLEJO.”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 34, septiembre de 2016; Tomo I; Pág. 517 1a. CCXXIX/2016 (10a.) Registro digital: 2012516

financieros); 2) una mala calidad de la atención o descuidos en el monitoreo del paciente; 3) la generación de lesiones al cuerpo del paciente que no eran justificadas (con base en el procedimiento que se debió seguir) o la utilización de materiales quirúrgicos inadecuados; o, 4) la contratación de personal poco capacitado para la realización del acto médico con el fin de reducir costo o manipular la facturación. Estos factores los podemos clasificar conforme la siguiente tabla:

Factores que generan errores médicos

Factores directos	Factores indirectos
• Capacitación y destrezas del médico • Delegación de responsabilidades en otros • Introducción de procedimientos nuevos • Exceso de confianza • Cuidados complejos/estancia prolongada • Ausencia de seguimiento • Inadecuada historia clínica • Falta de comunicación con el paciente y el equipo de salud • Errores de escritura y lectura	• Demora en la atención médica • Relaciones interpersonales deficientes (médicos-enfermeras-anestesiólogos) • Falta de recursos (humanos, materiales) • Desorden administrativo • Falta de estandarización en los procesos de atención • Incumplimiento de normas • Incumplimiento de procedimientos • <u>Aplicación de nuevas tecnologías</u> • Actitud del equipo de salud • Las cargas excesivas de trabajo • La fatiga y el estrés • Influencia del entorno

Fuente: El autor, con base en Sonia B. Fernández Cantón, Boletín CONAMED- OPS, marzo – abril 2016, México. Consúltese en: http://www.conamed.gob.mx/gobmx/boletin/pdf/boletin5/acto_medico.pdf

El médico debe de estar consciente de que su trabajo puede generar algún tipo de responsabilidad y el paciente debe entender cómo se desarrolla el trabajo del médico y hasta dónde llega su responsabilidad. En esta relación paciente-médico-acto médico-responsabilidad-pruebas, está presente a nivel mundial el acceso a la información y la transparencia en todo lo relacionado con padecimientos, diagnósticos, terapias, rehabilitación.

Hay ocasiones en las que el paciente no ve satisfecha sus expectativas, recurre a las diversas autoridades para exigir una indemnización o una reparación del daño ya sea por la vía civil o por la vía penal y en esa acción trae a su vez la necesidad de entender los hechos ocurridos, las evidencias (pruebas) y el derecho

que se aplica al caso concreto. En cuanto a los hechos ocurridos, el médico siempre debe cuidar diversas obligaciones que su propia profesión le impone.

La primera obligación del médico -quizás la de mayor importancia- está relacionada con la atención o el acto médicos y consiste en la correcta aplicación de las medidas que la *lex artis ad hoc* y la especialidad del médico contienen y que le dan el soporte para el diagnóstico y tratamiento. Aquí el médico es libre para decidir y hacer la prescripción cuidando esa *lex artis ad hoc*.

El médico que no cumpla esta obligación genera un acto ilícito. Y lo que se evalúa aquí es si la responsabilidad médica está dentro de esa libertad personal que le da la opción de decidir el tratamiento y demás actos médicos y aquellos otros factores que influyen o se presentan ya sea por la intervención del equipo médico (otros expertos invitados para intervenir en algún acto médico), los directivos y administrativos del establecimiento de salud y la propia persona jurídica con sus políticas, procedimientos, procesos y demás normas jurídicas y Éticas (programa de integridad corporativa, códigos de Ética, códigos de conducta, sistemas automatizados, manuales).

Un segundo tipo de obligación del médico está en la seguridad y el cuidado que deben de tener al realizar el acto médico y esto se relaciona con evitar siniestros relacionados con riesgos innecesarios, por el uso inadecuado de equipos o el mantenimiento inadecuado de equipos. Es obligación del médico, del personal que lo asiste y del establecimiento médico revisar permanentemente el estado del equipo que se utiliza para evitar accidentes.

Y finalmente, el tercer tipo de obligación tiene que ver con los resultados obtenidos o generados por su actuar. Aquí la comunidad internacional está de acuerdo en que no se puede obligar al médico a obtener resultados exitosos en su actuación. Pero, existen casos que contiene la legislación en los cuales sí se exige un tipo de resultado específico. Por ejemplo, cuando el personal de apoyo médico (enfermeras, anestesistas, laboratoristas, etcétera) cometen errores y se ve afectado el diagnóstico o el tratamiento. También en el caso del suministro de medicamentos adulterados, contaminados, caducos o alterados. Aquí también entran las prótesis o aditamentos que se pensaron para mejorar la vida del paciente

y que por algún descuido fallan impidiendo una buena rehabilitación o atención médica. También entra aquí los tratamientos o cirugías estéticas en los cuales el médico se compromete expresamente a cumplir con lo prometido.

En conjunto, estas tres obligaciones se conjugan para constituir lo que se llama el *Compliance* médico o de servicios de salud, que va de la mano con el *Compliance* farmacéutico y el *Compliance* tecnológico. El cumplimiento normativo mediante la prevención de riesgos dentro de las organizaciones está siendo utilizado en muchos hospitales privados que crean programas de *compliance* a través de los cuales se miden los riesgos legales de la organización.

¿Cómo saber qué tipo de responsabilidad existe por el incumplimiento de estas tres obligaciones? Para entender la responsabilidad médica tenemos que conocer lo que la ley mexicana entiende por actividades de atención médica (artículo 33 de la Ley General de Salud). Según la ley mexicana, existen cuatro tipos: 1) Preventivas, que incluyen las de promoción general y las de protección específica; 2) Curativas, que tienen como fin efectuar un diagnóstico temprano y proporcionar tratamiento oportuno; 3) De rehabilitación, que incluyen acciones tendientes a optimizar las capacidades y funciones de las personas con discapacidad, y 4) Paliativas, que incluyen el cuidado integral para preservar la calidad de vida del paciente, a través de la prevención, tratamiento y control del dolor, y otros síntomas físicos y emocionales por parte de un equipo profesional multidisciplinario.¹²¹

Otro concepto que debemos tomar en consideración para analizar la responsabilidad del médico lo encontramos en el término “equipo médico”. El artículo 262, fracción I de la *Ley General de Salud*, nos dice que por equipo médico se entiende: *“los aparatos, accesorios e instrumental para uso específico, destinados a la atención médica, quirúrgica o a procedimientos de exploración, diagnóstico, tratamiento y rehabilitación de pacientes, así como aquellos para efectuar actividades de investigación biomédica.”*

Vista esta definición, y haciendo una interpretación del contenido de la misma, se pudiera pensar que un software tradicional puede ser considerado como

¹²¹ Artículo 33 de la *Ley General de Salud*.

equipo médico ya que se utiliza o se encuentra dentro de un aparato (computadora, equipo de rayos X, equipo para la realización de tomografías, ultrasonido, análisis de ADN, etcétera) o en algún accesorio (teléfono móvil, tableta, impresora 3D) cuyo uso es específico (diagnóstico y/o tratamiento) y, también, hay aparatos y accesorios de IA cuyo su uso está enfocado para la investigación biomédica. Sin embargo, un sistema de IA basado en ML – DL o en cómputo cognitivo, como Watson Health, no entraría en esta categoría; o por lo menos, no es adecuado incorporarlo en esta categoría. Aquí vale la pena releer el caso del vuelo 447 de Air France y los argumentos expuestos previamente, sobre si Watson Health debe o no debe ser considerado como un dispositivo médico tradicional o como una pieza de software controlada por la voluntad del ser humano ya que funciona como un sistema interactivo que toma decisiones sin intervención del ser humano y supera los modelos de software que han sido certificados por distintas normas.

Esto se ha querido hacer con otros sistemas similares en industria tan sofisticadas como la aviación, certificando a los pilotos automáticos o directores de vuelo como una pieza más de software, pero en la práctica no funciona así. Madeleine Clare Elish, da el soporte a nuestro argumento, al estudiar diversos casos en los que el ser humano utiliza este tipo de tecnología y entra en una zona de “deformación moral” y de afectación a su capacidad de razonamiento y decisión, atribuyéndole una responsabilidad por una acción en la que no participó ni tenía posibilidades de tener el control de la decisión que tomó del sistema autónomo.¹²²

Por otro lado, existe la posibilidad de iniciar algún procedimiento de responsabilidad penal ya sea hacia la persona jurídica o hacia el médico responsable del acto. Aquí el dolo y la culpa están presentes para determinar la acción u omisión del sujeto. La responsabilidad médica se basa en el libre albedrío o en ese querer hacer u obrar para realizar alguna actividad considerada como delito y que debe ser analizada, entre otros elementos, mediante la valoración del dolo o la culpa.

¹²² Cfr. Clare Elish, Madeleine, “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction”, *Engaging Science, Technology, and Society*, volumen 5, 2019, pp. 41 y ss.

Obra dolosamente el que, conociendo los elementos del tipo penal, o previendo como posible el resultado típico, quiere o acepta la realización del hecho descrito por la ley, y obra culposamente el que produce el resultado típico, que no previó siendo previsible o previó confiando en que no se produciría, en virtud de la violación a un deber de cuidado, que debía y podía observar según las circunstancias y condiciones personales.¹²³

Las principales fuentes de los delitos culposos están relacionadas con el uso de vehículos, la industria de la construcción y la responsabilidad profesional del médico. De esta última fuente los penalistas han escrito una variedad de libros y artículos y han logrado construir teorías en torno a la causalidad, imputación objetiva, y al deber objetivo de cuidado. Además, existen diversos tipos penales relacionados con el acto médico y su responsabilidad penal. El Código Penal Federal contiene además algunos delitos relacionados con el actuar del médico en los artículos 228, 229, 230 (responsabilidad profesional); 246-IV (falsificación de documentos); 250 (usurpación de profesión); 288 y 302 (lesiones y homicidio); 335 (abandono de personas).

En México, Ruben Quintino intentó recopilar algunos casos relacionados con la *praxis médica* y el Derecho Penal (no solo imputación objetiva). En general, la problemática actual no ha llegado al punto de encontrar alguna afectación por el uso de sistemas inteligentes para la toma de decisiones médicas; no hay alguno que se relacione con el uso de sistemas inteligentes. El siguiente listado muestra ejemplos de problemáticas que se han generado por mala *praxis médica* y las consecuencias penales:

¹²³ Artículo 9 del Código Penal Federal.

Problemas de malas prácticas médicas y sus consecuencias

Problema	Consecuencias
Negó el consentimiento, por motivos religiosos, para operar a su hija. Fallece el paciente.	Se comete un delito por omisión de auxilio consumada instantáneamente donde intervinieron los coautores en forma de omisión simple dolosa
Médico hace una incisión cuya hemorragia no logró contener. Fallece el paciente.	Homicidio consumado instantáneamente donde intervino el autor directo en forma de acción culposa.
Víctima de un accidente de tránsito fallece por que el médico que lo trató no detectó a tiempo un edema cerebral causado	Homicidio consumado instantáneamente donde intervino el autor directo en forma de acción culposa
Una persona experimentó graves consecuencias después de haber interrumpido su tratamiento médico y fallece	La interrupción del tratamiento médico excluye la posición de garante del médico.
Un paciente aquejado de una grave enfermedad acude varias veces a servicios de urgencia en diferentes hospitales. Tanto el diagnóstico como el tratamiento resultaron contrarios a la lex artis y fallece	Hay una infracción de la debida diligencia cometida por los médicos tratantes. Homicidio consumado instantáneamente donde intervino el autor directo en forma de acción culposa.
El director de un hospital tomó con indiferencia conocer el grado académico de un practicante a quien convocó a realizar una cirugía. El paciente fallece a los pocos días. La persona no estaba titulada.	Homicidio consumado instantáneamente donde intervino el autor directo en forma de acción culposa, en concurso ideal con responsabilidad médica por ejercer la profesión sin título profesional consumada instantáneamente donde intervino el autor directo en forma de acción dolosa
Un médico practica una traqueotomía sin dominar suficientemente esa técnica. Fallece el paciente.	La acción del médico fue imprudente. Homicidio consumado instantáneamente donde intervino el autor directo en forma de acción culposa.

Fuente: El autor, con base en el libro *Responsabilidad penal del médico*, Editorial Arquinza, México, 2018.

Para darnos una idea de lo que implica una mala práctica médica y el tema de la responsabilidad del médico, según datos de la Organización Mundial de la Salud, los daños a la salud provocados por un acto médico o la alteración del estado del paciente producido por el médico (Iatrogenia, yatrogenia o iatropatogenia) traen como resultado:

- 2,6 millones de personas muertas cada año (daños que pudieron ser evitados en situaciones de atención médica adecuadas).
- Uno de cada diez pacientes sufre daños que pudieron ser evitados mientras reciben atención hospitalaria.
- Una de cada cuatro hospitalizaciones causa daños que pueden ser evitados a los pacientes.

- Cuatro de cada diez pacientes sufren daños que pueden ser evitados en los entornos de atención médica primaria y ambulatoria
- Las estimaciones muestran que hasta un 20% a 25% de la población general experimenta daños que pueden ser evitados durante la atención.¹²⁴

El error médico es algo que preocupa a la comunidad internacional. Errores en diagnóstico, errores en medicamentos, errores técnicos y, también, errores médicos por el uso de nuevas tecnologías. La iatropatogenia no solo proviene del médico, sino también de las instituciones y gobiernos por deficiencias del servicio, de la administración o la carencia de recursos. Las acciones imprudentes, impulsivas y no pensadas pueden causar un daño al paciente. La iatropatogenia puede ser causada por acción o por omisión y se da en cuestiones quirúrgicas, al recetar medicamentos, al emitir un diagnóstico, al interpretar un estudio, por un conjunto de acciones u omisiones relacionadas con dos o más de los anteriores supuestos y, finalmente, la iatropatogenia por uso de equipo o tecnología novedosa.¹²⁵

Los errores de diagnóstico tienen orígenes tanto cognitivos como sistémicos.

H. Singh, LA Petersen y EJ. Tomas nos explican lo siguiente:

El enfoque de sistemas atribuye errores de diagnóstico a fallas en la entrega o función, como el seguimiento de resultados de laboratorio e imágenes anormales o críticos, presiones de productividad, atención discontinua y transferencias, procesos y políticas poco estandarizados, sistemas de comunicación deficientes dentro de las organizaciones y falta de información esencial relevante para el paciente. Incluso las personas bien capacitadas corren el riesgo de cometer errores graves al trabajar en sistemas mal diseñados. Por lo tanto, se ha sugerido el rediseño de los sistemas para facilitar los procesos de toma de decisiones clínicas de los médicos y proporcionarles un entorno de trabajo a prueba de fallas y una estructura organizativa confiable para minimizar los errores de diagnóstico. Un segundo enfoque proviene de la psicología cognitiva: el estudio de cómo los médicos razonan, formulan juicios y toman decisiones diagnósticas. El proceso de elaboración de un diagnóstico consiste en la recopilación de datos, la síntesis de datos y la verificación de datos, y cada una de estas etapas puede ser propensa a errores. Por lo tanto, los errores de diagnóstico pueden surgir de fallas en la percepción, heurísticas fallidas (atajos en el razonamiento o “reglas generales”), sesgos y otros estados mentales (como el afecto) que se denominan disposiciones cognitivas o afectivas para responder (CDR o ADR). Este enfoque enfatiza que

¹²⁴ https://www3.paho.org/hq/index.php?option=com_content&view=article&id=15372:world-patient-safety-day-2019&Itemid=39594&lang=es

¹²⁵ María Angulo, Eél, “Errores médicos, las sentencias de muerte de 2,6 millones de pacientes por año” en *France24*, del 15 de septiembre de 2019, consulte en: <https://www.france24.com/es/20190914-errores-medicos-oms-muerte-salud>

podemos reducir los errores de diagnóstico diseñando estrategias basadas en el análisis de CDR y ADR. En un estudio reciente de 93 errores de diagnóstico de Graber et al, se encontraron 15,548 factores cognitivos o relacionados con el sistema diferentes (5,9 por caso). Los factores relacionados con el sistema contribuyeron al error diagnóstico en el 65% de los casos y los factores cognitivos en el 74%, destacando el origen multifactorial de dichos errores.¹²⁶

Para concluir esta sección, nos debemos preguntar también: ¿qué puede considerarse o qué puede generarnos algún indicio de afectación a Derechos Humanos y cómo se debe calcular la reparación del daño? Por cierto, en este tema debemos partir de un parámetro de control de la responsabilidad diferente para generar la reparación del daño. Uno distinto al que se utiliza para calcular una indemnización o una sanción penal o civil; y debe tomar -según sea el caso- la situación, pasada, presente y futura de la víctima, que podemos denominar como una reparación integral y justa derivada de una violación de Derechos Humanos.

La explicación que damos a continuación se basa en la evolución que ha tenido el Derecho Constitucional Mexicano de 1917 a la fecha. México, como otros países latinoamericanos, ha venido modificando su constitución en diversos temas relacionados con el respeto de los Derechos Humanos de las personas, pero existía una particularidad, históricamente la Constitución mexicana no hablaba de Derechos Humanos sino de garantías individuales. Este término, heredado y utilizado por lo menos desde 1857 hasta antes de la entrada en vigor de la reforma constitucional de Derechos Humanos (junio de 2011) acotaba los Derechos Humanos a los 29 artículos y a su contenido y se generaba polémica respecto el tratamiento de otros derechos y su protección, como, por ejemplo, los derechos político-electorales o los derechos laborales.

Este tipo de restricciones también nos lleva al ámbito constitucional y su relación con el concepto reparación del daño por violación de Derechos Humanos. Nuestro sistema jurídico no contemplaba esta posibilidad ya que no tenían vida dentro del texto de la Constitución y las indemnizaciones o reparación del daño estaban acotadas a la teoría de la responsabilidad basada en el hecho ilícito ya sea

¹²⁶ Singh H, Petersen LA, Thomas EJ., Understanding diagnostic errors in medicine: a lesson from aviation, Qual Saf Health Care, 2006, junio 15, pp. 159-64. doi: 10.1136/qshc.2005.016444. PMID: 16751463; PMCID: PMC2464840. <https://pubmed.ncbi.nlm.nih.gov/16751463/> El caso que se utiliza es tomado de este texto. [20 de julio de 2022]

por incumplimiento contractual, extracontractual o por la comisión de algún delito. En materia administrativa se encontraba regulado esto con reglas similares añadiendo el tipo de sujeto activo basado en la actividad del Estado, es decir, en que fuera servidor público o una institución pública.

Al momento de incluir el concepto Derechos Humanos y el principio pro persona y generar doctrina y jurisprudencia para entender sus efectos y consecuencias dentro del sistema jurídico mexicano, se llega a la conclusión que no es lo mismo analizar violaciones a Derechos Humanos que afectaciones a las personas por hechos ilícitos. Tampoco es igual si estas violaciones son hechas por un particular o por algún agente del gobierno.

Tampoco se puede equiparar una violación de Derechos Humanos con una acción de responsabilidad civil ya que esta última es un concepto más reducido. El derecho a una indemnización justa o reparación del daño integral está constituido por un conjunto de elementos que la Primera Sala de la Suprema Corte de Justicia de la Nación ha venido determinado a partir de 2018, con base en el propio concepto Derechos Humanos. Las afectaciones generadas de la víctima y sus consecuencias en el proyecto de vida -pasado, presente y futuro- y que se vinculan con el concepto reparación integral del daño por violaciones a Derechos Humanos. Por la relevancia de la tesis de jurisprudencia, consideramos necesario su incorporación literal:

En la Constitución Política de los Estados Unidos Mexicanos, desde su promulgación en 1917 y hasta el 2000, no existía noción de “reparación del daño”, sino que su regulación se realizó en la legislación secundaria. Esta situación cambió paulatinamente con las siguientes cuatro reformas constitucionales: (1) la de 21 de septiembre de 2000 que introdujo en el texto del artículo 20 constitucional, un apartado B, en el que se estableció un elenco mínimo de derechos de las víctimas u ofendidos por la comisión de delitos, uno de los cuales fue la facultad de solicitar una reparación del daño; (2) la de 14 de junio de 2002 que reformó el artículo 113 constitucional para adicionarle un segundo párrafo (que actualmente se encuentra en la parte final del artículo 109), para establecer que la responsabilidad del Estado por su actividad administrativa irregular es objetiva y directa, y da lugar al pago de una indemnización a favor de la persona que haya resentido el daño; (3) la de 18 de junio de 2008 en materia procesal penal que trasladó el catálogo de derechos de las víctimas y ofendidos al apartado C del artículo 20 constitucional, e incorporó en su fracción VII, el derecho a impugnar determinaciones del Ministerio Público que afecten la obtención de la reparación del daño; y (4) la de 29 de julio de 2010 que introdujo en la Constitución Federal el fundamento de las acciones colectivas, dejando a la legislación secundaria la regulación de los mecanismos de reparación del daño. Ante estas reformas, la legislación secundaria desarrolló el contenido de las reparaciones o de la indemnización bajo una base eminentemente civil y con un contenido apoyado principalmente en la teoría de las obligaciones. Esta situación cambió

con la reforma constitucional en esta materia, publicada el 10 de junio de 2011, la cual incluyó en el tercer párrafo de su artículo 1o. un catálogo de las obligaciones genéricas y los deberes específicos del Estado mexicano en materia de Derechos Humanos, dentro de los cuales se reconoció la “reparación por violaciones a Derechos Humanos”. Al respecto, el deber de “reparar” tales violaciones no fue incluido en el dictamen original de reforma elaborado por las Comisiones Unidas de Puntos Constitucionales y de Derechos Humanos de la Cámara de Diputados, sino que surgió hasta el dictamen suscrito el 7 de abril de 2010 por las Comisiones Unidas de Puntos Constitucionales y de Estudios Legislativos de la Cámara de Senadores, con opinión de la Comisión de Reforma del Estado, sin que la adición fuese objeto de cambios durante el resto del proceso de reforma constitucional. Para entender el concepto de “reparación” incorporado a la Constitución, es importante señalar que el Senado invocó el concepto de “reparación integral” desarrollado en el marco de las Naciones Unidas, partiendo de los “principios y directrices básicos sobre el derecho de las víctimas de violaciones manifiestas de las normas internacionales de Derechos Humanos y de violaciones graves del derecho internacional humanitario a interponer recursos y obtener reparaciones”.¹²⁷

Este cambio de paradigma aún se mantiene aislado y sin ser entendido por muchos abogados y jueces. En general, la forma de litigar y resolver los asuntos se mantiene sin este criterio y tal como se ha venido argumentando desde antes del cambio al artículo primero constitucional (2011). Para complementar la tesis anterior se hizo una búsqueda de otros criterios, encontrándose varios que toman ya este soporte conceptual y lo van ajustando a casos específicos. Los principales temas y argumentos emitidos por la Suprema Corte de Justicia de la Nación (SCJN) relacionados con la violación a Derechos Humanos y su justa indemnización o reparación integral son las siguientes:

1. El impacto del acto lesivo en la reparación del daño y el derecho a una indemnización justa van de la mano. - Para reparar un daño, dice la SCJN, es importante poner énfasis en el “impacto multidimensional del hecho lesivo, incluyendo tanto el sufrimiento de la víctima como la cadena de impactos negativos desatada por aquél, sin que esto se aplique a casos derivados de responsabilidad contractual y los que generen daños meramente patrimoniales, pues aquí los efectos son más bien unidimensionales.” De esta forma, bajo la figura de indemnización justa, es obligatorio revisar si los

¹²⁷ “REPARACIÓN INTEGRAL DEL DAÑO POR VIOLACIONES A DERECHOS HUMANOS. ORIGEN DE SU INCORPORACIÓN AL TEXTO CONSTITUCIONAL EN LA REFORMA DE 10 DE JUNIO DE 2011”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 400 1a. CCCXXXVII/2018 (10a.) Registro digital: 2018805

montos cubren todas las afectaciones y consecuencias (patrimoniales y extrapatrimoniales), derivadas de un hecho ilícito, para redignificar y rehabilitar a las personas. La Primera Sala considera que la justa indemnización tiene como primera finalidad, replantear los alcances de los procedimientos estrictamente indemnizatorios –como los juicios civiles por responsabilidad extracontractual o los de responsabilidad patrimonial– para garantizar las compensaciones dictadas. Por tanto, ya no tiene tanta validez el uso de topes o límites a los montos indemnizatorios por considerarse contrarios al derecho a la justa indemnización.¹²⁸

2. Las diferencias entre violaciones de Derechos Humanos y hechos ilícitos influyen para el cálculo de una indemnización justa.- La SCJN delimita algo fundamental: “No es lo mismo analizar violaciones a Derechos Humanos que hechos ilícitos en general, y también existen diferencias dependiendo de si el estudio se realiza en sede administrativa –jurisdiccional o cuasi-jurisdiccional– o en una acción de responsabilidad civil o en amparo, pues cada vía admite cierto tipo de medidas reparadoras y tiene reglas para determinar su procedencia”. Por lo que hace a la individualización de indemnizaciones, para que se consideren “justas o integrales”¹²⁹, deben comprender porcentajes y objetivos diversos. Para que una indemnización sea considerada justa cuando debe cumplir dos requisitos: 1) la reparación integral del daño y 2) la individualización de la condena según cada caso. La

¹²⁸ “DERECHO A UNA JUSTA INDEMNIZACIÓN. PARA DETERMINAR EL MONTO INDEMNIZATORIO, SE DEBE ATENDER A LA MULTIPLICIDAD DE CONSECUENCIAS DEL HECHO ILÍCITO”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 290 1a. CLXXXVII/2018 (10a.) Registro digital: 2018644

¹²⁹ La Primera Sala de la SCJN también ha manifestado que el derecho fundamental a una reparación integral o justa es un derecho sustantivo cuya extensión debe tutelarse en favor de los gobernados sin restricciones innecesarias. La SCJN toma los criterios emitidos por la Corte Interamericana de Derechos Humanos y entiende que el derecho a la reparación integral permite anular todas las consecuencias del acto ilícito y restablecer la situación que debió haber existido con toda probabilidad, si el acto no se hubiera cometido, y de no ser esto posible, procede el pago de una indemnización justa como medida resarcitoria por los daños ocasionados, lo cual no debe generar una ganancia a la víctima, sino que se le otorgue un resarcimiento adecuado. Una indemnización es injusta -dice la SCJN- cuando se le limita con topes o tarifas, “y en lugar de ser el Juez quien la cuantifique justa y equitativamente con base en criterios de razonabilidad, al ser quien conoce las particularidades del caso”. “DERECHO FUNDAMENTAL A UNA REPARACIÓN INTEGRAL O JUSTA INDEMNIZACIÓN. SU CONCEPTO Y ALCANCE.”. [J]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 41, abril de 2017; Tomo I; Pág. 752 1a./J. 31/2017 (10a.) Registro: 2014098.

SCJN recomienda cuidar: “i) la naturaleza y extensión de los daños causados, esto es, si son físicos, mentales o psicoemocionales; ii) la posibilidad de rehabilitación de la persona afectada; iii) la pérdida de oportunidades, en particular las de empleo, educación y prestaciones sociales; iv) los daños materiales, incluidos los ingresos y el lucro cesante; v) los perjuicios inmateriales; vi) los gastos de asistencia jurídica o de expertos, medicamentos y servicios médicos, psicológicos y sociales; vii) el nivel o grado de responsabilidad de las partes; viii) su situación económica; y ix) demás características particulares”. ¿se pueden modificar los procedimientos tradicionales y métodos de cálculo de la indemnización? La SCJN advierte que solo procede la modificación cuando las reglas tradicionales de la indemnización no resulten compatibles con el derecho a obtener una justa indemnización; es decir, en caso de ser compatibles, se deben seguir.¹³⁰

3. Inconstitucionalidad de fijar topes máximos y mínimos en la reparación del daño por violaciones a Derechos Humanos. - Al analizar la reparación integral del daño en casos que impliquen violaciones a Derechos Humanos se debe tener presente siempre el impacto multidimensional de un hecho lesivo, incluyendo tanto el sufrimiento de la víctima como la cadena de impactos en otros Derechos Humanos. Las indemnizaciones serán consideradas justas cuando su cálculo se realice con base en el encuentro de dos principios: “el de reparación integral del daño y el de individualización de la condena según las particularidades de cada caso, incluyendo: 1) la extensión de los daños causados y su naturaleza (físicos, mentales o psicoemocionales); 2) la posibilidad de rehabilitación; 3) la pérdida de oportunidades, en particular las de empleo, educación y prestaciones sociales; 4) los daños materiales (ingresos y el lucro cesante); 5) los daños inmateriales; 6) los gastos de asistencia jurídica o de expertos, medicamentos y servicios médicos, psicológicos y sociales; 7) el nivel o grado de responsabilidad de las partes; 8) su situación económica; y, 9) las

¹³⁰ “DERECHO A UNA JUSTA INDEMNIZACIÓN. PARÁMETROS QUE RIGEN LA INDIVIDUALIZACIÓN DE LOS MONTOS QUE LA INTEGRAN”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 292 1a. CXC/2018 (10a.) Registro digital: 2018645

demás características particulares. Por ello, el derecho a la reparación integral del daño en casos que afecten Derechos Humanos es incompatible con la existencia de topes, tarifas o montos máximos y mínimos que impidan que la cuantificación de una indemnización atienda a las características específicas de cada caso”.¹³¹

La SCJN crea una excepción o una prevención que se debe de tomar en cuenta por el juzgador y que tiene que ver con el cumplimiento de las reglas que rigen la reparación del daño y lo que la doctrina pensó desde sus orígenes. Es decir, en el sistema jurídico mexicano hay hechos ilícitos ya regulados por el Derecho Civil y el Derecho Administrativo que pueden ser considerados como típicos, al estar su solución dentro del Código Civil o la ley administrativa y ante los cuales las personas pueden acudir a los mecanismos jurisdiccionales diseñados para su atención (hay un marco legislativo que los regula). Y hay otros que son atípicos o violaciones a Derechos Humanos (hechos victimizantes), que deben atenderse desde el marco constitucional aplicable y revisando la forma en la cual el ordenamiento jurídico puede contribuir al respecto.

Los asuntos de negligencia médica pueden encuadrar en ambos supuestos, sin embargo, se debe tomar, en primera instancia, el camino del marco legislativo que regula, por ejemplo, la responsabilidad patrimonial del Estado y la responsabilidad civil extracontractual, con independencia de que en el fondo puedan estar en juego los derechos de la vida y la integridad. “Las normas y los procedimientos que formen parte de las instituciones respectivas deberán interpretarse de conformidad y contrastarse con el parámetro de control de regularidad, de modo que en cada caso se tutelen adecuadamente los derechos de las personas involucradas.”.¹³²

¹³¹ “REPARACIÓN INTEGRAL DEL DAÑO POR VIOLACIONES A DERECHOS HUMANOS. TOPES MÍNIMOS Y MÁXIMOS DE LA CUANTIFICACIÓN DE LAS INDEMNIZACIONES, SU INCONSTITUCIONALIDAD”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 402 1a. CXCV/2018 (10a.) Registro digital: 2018806

¹³² “TRANSVERSALIDAD DE LOS DERECHOS HUMANOS. EL QUE UN HECHO PUEDA CALIFICARSE COMO VICTIMIZANTE POR CONLLEVAR VIOLACIONES A AQUÉLLOS, NO IMPLICA HACER A UN LADO LAS REGLAS QUE RIGEN A LAS INSTITUCIONES DISEÑADAS PARA EXIGIR LA REPARACIÓN DEL DAÑO CORRESPONDIENTE, ASÍ COMO A LA DOCTRINA

Para finalizar este apartado, es importante mencionar que la Suprema Corte de Justicia de la Nación, en 2016, emitió diversas jurisprudencias sobre la responsabilidad del médico (contractual y extracontractual), sobre el acto médico, la negligencia y la responsabilidad de las instituciones hospitalarias. En general, se puede decir que en todas está presente la 1) valoración de la conducta doloso o culposa (error doloso y culposos), 2) la obligación que tienen los médicos y las instituciones de salud de brindar sus servicios bajo la premisa de realizar todas las conductas necesarias para la consecución de los fines por los cuales fueron contratados siguiendo la *Lex Artis ah hoc*; 3) sancionándose la falta de cumplimiento o acatamiento de ésta. A continuación, expondremos una nueva categoría de responsabilidad jurídica (no vigente) que se denomina responsabilidad algorítmica.

B) Categorías de la Responsabilidad algorítmica y otros conceptos

En esta segunda dimensión explicaremos los dos conceptos que nos llevan a la creación de una nueva categoría de responsabilidad jurídica y que hemos denominado responsabilidad algorítmica, la cual, surge o es producida por el llamado Agente Artificial Moral Autónomo (AAMA). Esta dimensión tiene su origen, indirectamente, en la imposibilidad que existe en la primera dimensión de atender con precisión los casos de uso de IA que toma decisiones sin intervención humana y también tiene un soporte en las aportaciones conceptuales de la Ética ya que utilizamos conceptos de dicha disciplina (en particular, agencia y agencia moral). Recordemos que en la primera dimensión la voluntad de la persona y las consecuencias de sus actos o los riesgos que se generan por el uso de artefactos peligrosos son la fórmula de la responsabilidad jurídica. Sin embargo, esta forma de categorizar la responsabilidad no contempla el uso de sistemas autónomos que toman decisiones sin intervención de los seres humanos, es decir, funciona sin el acto volitivo de la persona.

QUE SE HA ELABORADO EN TORNO A ÉSTA". [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 464 1a. CLXXXVIII/2018 (10a.) Registro digital: 2018862

Para ello nos debemos preguntar si los seres humanos son los únicos que pueden ser responsables del comportamiento moral de las máquinas (o de la tecnología) y si el puente que existe entre el Derecho y la Ética te lleva a un nuevo planteamiento. Siguiendo las categorías jurídicas vigentes de responsabilidad, evidentemente que la respuesta es que el ser humano es el único responsable del comportamiento - moral y jurídico- de las máquinas. Solo los seres humanos y nadie más puede ser responsable ante la norma jurídica. Así fue hecho nuestro sistema porque no existían máquinas que funcionaran de manera autónoma del ser humano, ni en la época de Kant, ni en la de Hart o Dworkin. Y en ese sentido, nuestro sistema de responsabilidad funcionaba bien; cimentado en el acto volitivo o participación de las personas (directa o indirectamente hay una participación del ser humano o un deber de cuidado del padre, tutor o patrón) para definir si son acreedores de una sanción o castigo por un hecho ilícito o por un delito.

Para poner el contexto, se toman como ejemplo las reflexiones del autor Richard De George¹³³, hechas en su libro *The Ethics of Information Technology and Business* (2003), en el cual se explica que los desafíos éticos en los negocios son cambiantes y muy complejos en la medida que las prácticas comerciales se reconfiguran como resultado de la informática y las tecnologías de la información; y crea el Mito de la Computación Amoral y de la Tecnología de la Información (MACIT). Este mito -afirma- nos ciega a los poderosos cambios que están teniendo lugar como resultado de la informática y la tecnología de la información y consiste en lo siguiente: “Aquellos que construyen, programan, ejecutan, poseen y/o administran las computadoras o sistemas de información son los únicos que pueden ser considerados moralmente responsables de los resultados”.¹³⁴

La hipótesis de De George ha sido analizada por diversos académicos, entre ellos Deborah G. Johnson, quien en 2014 publica un artículo titulado: “Technology

¹³³ Es considerado como el pionero en el uso de la Ética para los negocios y la tecnología de la información.

¹³⁴ De George, R. T., *The ethics of information technology and business*, Malden, Blackwell Publishing, 2003. De George habla de privacidad, del comercio electrónico, de la propiedad intelectual, de los datos, los dispositivos electrónicos y papel del trabajo humano bajo estas nuevas tecnologías (así como el grado en que son monitoreadas las personas).

with No Human Responsibility?”¹³⁵, en el cual hace una serie de reflexiones para explicar por qué la hipótesis de De George ya no tiene vigencia a raíz de las nuevas tecnologías autónomas (v.gr. Redes Neuronales Profundas). Johnson, explica como el modelo de responsabilidad jurídica tradicional puede quedar limitado con estas tecnologías al grado de no ubicar qué ser humano es el responsable ante la ley (o por lo menos, no está tan claro cómo hacerlo y en qué momento surge su responsabilidad).

La tecnología no solo tiene ingredientes de moralidad, sino también existe moralidad en su creación y utilización y actualmente se discute cómo se están creando nuevos conceptos como el de *Machine Learning – Deep Learning*, datos, opacidad, sesgos algorítmicos interpretabilidad y explicabilidad, y la posibilidad de utilizarlos para construir el concepto de Agente Artificial Moral Autónomo (AAMA), que a su vez permitiría dividir la responsabilidad que le corresponde a la máquina (responsabilidad algorítmica) de la responsabilidad que les corresponde a los seres humanos. Es por ello por lo que decidimos proponer la adición al sistema jurídico de los conceptos Agente Artificial Moral Autónomo (AAMA) y responsabilidad algorítmica.

Primer concepto. – Agente Artificial Moral Autónomo. Para entender este concepto, debemos recordar qué es la agencia y/o el agente moral, conceptos que por siglos han sido estudiados. En principio decir que estos conceptos no fueron articulados pensando en la posibilidad de que un objeto (sistema o máquina) diera indicaciones a un ser humano y que éste tomará decisiones siguiendo a la máquina y no a su conciencia o con su inteligencia. Tampoco era posible que la norma jurídica regulara a la máquina y sus decisiones, al momento de generar una afectación a la persona, a su comunidad o grupo o a toda la humanidad en conjunto. Sin embargo, el Derecho se debe acercar a la Filosofía para analizar estos temas e incorporarlos como una nueva categoría con base en el análisis de los orígenes, estructura y funciones de los sistemas inteligentes. Construir una categoría

¹³⁵ Cfr. Johnson, D.G., Technology with No Human Responsibility?, *J Bus Ethics* 127, 2015, pp. 707-715, <https://doi.org/10.1007/s10551-014-2180-1> (20 de julio de 2022)

adicional de responsabilidad con conceptos que se configuren y complementen con el modelo tradicional de responsabilidad vigente.

En estos temas hay dos posturas teóricas. Aquellos que están a favor de que las máquinas y/o sistemas algorítmicos y/o sistemas inteligentes puedan ser considerados como un agente moral (previa valoración de su estructura de IA); y otros, que se niegan a esa posibilidad ya que la Moral fue pensada únicamente para seres humanos. Sin influir sobre qué postura tomar, sí creemos que el tema debe ser debatido por los juristas que materialicen esta segunda dimensión en norma jurídica y que entiendan su relevancia para la construcción de nuevas categorías jurídicas que permitan la asignación de la responsabilidad a ciertos sistemas autónomos. Johnson abre al debate su definición de agente artificial:

“...aquellos dispositivos computacionales que *realizan tareas en nombre de los seres humanos y lo hacen sin control o intervención humana inmediata y directa*. Algunos agentes artificiales son programas de software, por ejemplo, bots que realizan búsquedas en internet web; estos programas son puramente computacionales. Otros agentes artificiales son las combinaciones de hardware y software, por ejemplo, robots; estas combinaciones tienen componentes computacionales de toma de decisiones incrustados en sus estructuras incorporadas. Algunos argumentan que debido a que ciertos agentes artificiales aprenden mientras operan, aquellos que diseñaron o desplegaron esos agentes pueden no ser capaces de controlar o incluso predecir lo que harán sus agentes. *A medida que estos agentes se vuelven cada vez más autónomos, dice el argumento, ningún ser humano será responsable de su comportamiento.*”¹³⁶

De su definición, dos son los elementos que nos obligan a crear una nueva categoría de responsabilidad jurídica: El primero tiene que ver con la naturaleza de los dispositivos, es decir, son máquinas que realizan tareas de los seres humanos sin su control o intervención directa; y, el segundo, que estos agentes cada vez se vuelven más autónomos y llegará algún momento en que ningún ser humano sea responsable de su comportamiento. Es por ello que los consideramos la fuente de la responsabilidad algorítmica.

¹³⁶ “The term artificial agent refers broadly to computational devices that perform tasks on behalf of humans and do so without immediate, direct human control or intervention. Some artificial agents are software programs, e.g., bots that perform Internet web searches; these programs are purely computational. Other artificial agents are hardware–software combinations, e.g., robots; these combinations have computational decision-making component embedded in their embodied structures. Some argue that because certain artificial agents learn as they operate, those who designed or deployed those agents may not be able to control or even predict what their agents will do. As these agents become increasingly more autonomous, the argument goes, no humans will be responsible for their behavior.” Cfr. Johnson, D.G. Technology with No Human Responsibility?. *J Bus Ethics*, 127, pp. 707-715, 2015. <https://doi.org/10.1007/s10551-014-2180-1> (20 de julio de 2022)

El comportamiento de estas máquinas no surge de un ser humano programador que en la etapa de producción fija las reglas de su comportamiento, sino que la máquina aprende a aprender y a tomar decisiones y puede corregir y mejorar durante su funcionamiento cotidiano con el solo hecho de tener contacto con los datos de la red a la que se conecte o del sensor que le genere datos del ambiente.

Consideramos que Johnson en sus argumentos logra disprobar lo que De George sostenía en 2003 que los humanos, no la tecnología, siempre son responsables de lo que se hace con la tecnología y esto nos lleva la actualización de nuestras normas jurídicas.

Partamos de un hecho real: La IA, siempre ha buscado crear entes artificiales capaces de simular la mente y el razonamiento humano. Y desde un lustro atrás, existen sistemas inteligentes con capacidades nunca vistas, muy similares a las de los seres humanos. Estos sistemas pueden hacer tareas altamente complejas (como lo hemos explicado a lo largo de esta investigación). Esto no debe quedar sin regulación, ni tampoco sin supervisión y sin control.

Sobre todo, en el campo de la Ética de las máquinas, hay propuestas para dotar a los agentes autónomos de “mecanismos éticos para proporcionarles la capacidad de abordar los problemas que pueden surgir en la relación entre los agentes autónomos y los seres humanos, tales como dilemas morales donde la acción o inacción de un agente autónomo puede llevar a un ser humano a sufrir un daño”¹³⁷.

Pensemos una cosa, al momento en que el ser humano decide apoyarse en las máquinas para ciertas actividades, se genera una delegación de decisiones que cambian el mundo (como cualquier agencia). Los expertos que apoyan esto se preguntan ¿qué capacidades morales debe tener un sistema para considerarlo

¹³⁷ Cfr. Cervantes, JA., López, S., Rodríguez, LF. et al., Artificial Moral Agents: A Survey of the Current Status, *Sci Eng Ethics*, 26, 2000, pp. 501–532. <https://doi.org/10.1007/s11948-019-00151-x> Los autores explican: “The objective of implementing adjustable autonomy in AAs is to maintain human’s global control over AAs in order to avoid undesirable or inappropriate autonomous behaviors (Mostafa et al. 2019). In addition, in the field of machine ethics, a crucial challenge has been to endow AAs with ethical mechanisms in order to provide them with the ability to address issues that may arise in the relationship between AAs and human beings, such as moral dilemmas where the action or inaction of an AA may lead a human to suffer harm”

como un AAMA? ¿cómo deben modelarse sus capacidades Éticas y cómo deben ser implementadas? Las respuestas están en el respeto a los Derechos Humanos y en los distintos tipos de Ética que se tiene (sin contar el tipo de cultura o sociedad a la que se aplique). Si el sistema tiene estas capacidades mentales y han sido construidas con IA basada en Machine Learning – Deep Learning, debemos aceptar que sí encuadra en un AAMA.

José-Antonio Cervantes, Sonia López, Luis-Felipe Rodríguez, Salvador Cervantes, Francisco Cervantes y Félix Ramos, un equipo de investigadores mexicanos, decidieron hacer una taxonomía que busca mostrar el estado del arte del tema de los agentes morales artificiales. El documento, novedoso y único en su tipo, permite entender el tema desde el ámbito de la Ingeniería. Los investigadores concluyen lo siguiente:

“los seres humanos buscan delegar parte de su poder de toma de decisiones en los AA, aumentando así el alcance de sus actividades. Dotar a los AA de mecanismos éticos ha sido un enfoque inicial en el campo de la Ética de la máquina que intenta lidiar con algunos problemas probables que pueden surgir en la relación entre los AA y los seres humanos y coexistir de una manera segura, en armonía y con confianza... Como resultado de esta revisión, **se propuso una nueva taxonomía que tiene como objetivo ser útil para comprender las ventajas y limitaciones de las AMA basadas en la estrategia de diseño (es decir, de arriba hacia abajo, de abajo hacia arriba, híbrida e implícita) y los criterios (es decir, Ética normativa, situacionismo, código empírico y no malicioso) empleados para tomar decisiones Éticas.** De la revisión presentada, **es evidente que actualmente no existen sistemas generales de inteligencia artificial capaces de tomar decisiones morales sofisticadas como lo hacen los humanos. Sin embargo, hay algunos prototipos que tratan con ciertas cuestiones morales**, pero estos prototipos todavía están en sus primeras etapas de desarrollo. Todos estos prototipos han sido probados utilizando casos básicos en los que el entorno de prueba está bien definido y controlado, en contraste con escenarios reales donde situaciones morales críticas o inciertas podrían surgir inesperadamente... **Este trabajo concluye que debido a la interacción actual entre los sistemas AAs y los seres humanos, los mecanismos éticos para AMAs no son una opción potencial para los futuros sistemas autónomos, pero una necesidad real de tratar...**”¹³⁸

¹³⁸ *Ibidem*. Los autores dicen: “As described in this paper, human beings are looking to delegate part of their decision- making power to AAs, thus increasing the scope of their activities. Endowing AAs with ethical mechanisms has been an initial approach in the field of machine ethics that attempts to deal with some likely issues that may arise in the relationship between AAs and human beings and coexist in a safe way, in harmony, and with confidence. This paper presented an analysis of the current status of ethical artificial agents. As a result of this review, a new taxonomy was proposed that aims to be useful for understanding advantages and limitations of AMAs based on the design strategy (i.e., top-down, bottom-up, hybrid, and implicit) and criteria (i.e., normative ethics,

Para llegar a esta conclusión, los académicos estudian la clasificación de agentes éticos propuesta por J. H. Moor¹³⁹, que a su vez retoma Carissa Veliz en su artículo *Moral Zombies: why algorithms are not moral agents*. Los cuatro tipos de robots éticos, vistos como AAMA, son:

1) ethical impact agents (agents whose actions have ethical consequences; most or all autonomous systems count as ethical impact agents); (2) implicit ethical agents (those that have been designed with ethics in mind, e.g. ATMs); (3) explicit ethical agents (agents that can act from ethics, and not merely in accordance with ethics; they can identify ethical problems and work out solutions for themselves); (4) full ethical agents (they can do what explicit ethical agents do and, in addition, they have consciousness, intentionality, and free will). My claim in this paper is that explicit ethical agents and full ethical agents are one and the same category. In other words, I claim we will not get an agent that can identify ethical problems and respond adequately to them without sentience. Given that my argument pivots on sentience, the most interesting perspective (and contrary to mine) is that of Luciano Floridi and J.W. Sanders, who not only argue that artificial autonomous agents can be considered moral agents, but also that they constitute an instance of 'mind-less morality' (2004, 351). In disagreement with Floridi and Sanders, I contend that sentience is necessary for moral agency, because conscious experience of the kind that allows for feelings like pleasure, pain, and empathy seems necessary for moral agency.¹⁴⁰

situationism, empirical, and non-malicious code) employed to make ethical decisions. From the review presented, it is evident that currently there are no general artificial intelligence systems capable of making sophisticated moral decisions as humans do. However, there are a few prototypes that deal with certain moral issues, but these prototypes are still in their early stages of development. All these prototypes have been tested using basic cases where the test environment is well-defined and controlled, in contrast to real scenarios where critical or uncertain moral situations could arise unexpectedly. Moreover, the review illustrates that there is a long way to go (from a technological perspective) before this type of artificial agent can replace human judgment in difficult, surprising or ambiguous moral situations. This paper concludes that because of the current interaction between AAs systems and humans, ethical mechanisms for AMAs are not a potential option for future autonomous systems, but an actual need to deal with. Perhaps their ethics could be different from human ethics, but at this moment, models of human ethics are the guides most used by researchers to develop AMAs."

¹³⁹ J. H. Moor, "The nature, importance, and difficulty of machine ethics", *IEEE Intelligent Systems*, 21, 4, pp. 18-21, 2006, y "Four Kinds of Ethical Robots". *Philos Now* 72, pp. 12-14.

¹⁴⁰ Cfr. Carissa Véliz, "Moral zombies: why algorithms are not moral agents", *AI & Soc* 36, pp. 487-497, 2021. <https://doi.org/10.1007/s00146-021-01189-x> (1) agentes de *impacto ético* (agentes cuyas acciones tienen consecuencias Éticas; la mayoría o todos los sistemas autónomos cuentan como agentes de impacto ético); (2) *agentes éticos implícitos* (aquellos que han sido diseñados con la Ética en mente, por ejemplo, cajeros automáticos); (3) *agentes éticos explícitos* (agentes que pueden actuar desde la Ética, y no simplemente de acuerdo con la Ética; pueden identificar problemas éticos y encontrar soluciones por sí mismos); (4) *agentes éticos plenos* (pueden hacer lo que hacen los agentes éticos explícitos y, además, tienen conciencia, intencionalidad y libre albedrío). Mi afirmación en este documento es que los agentes éticos explícitos y los agentes éticos completos son una y la misma categoría. En otras palabras, afirmo que no tendremos un agente que pueda identificar problemas éticos y responder adecuadamente a ellos sin sensibilidad. Dado que mi argumento gira en la sensibilidad, la perspectiva más interesante (y contraria a la mía) es la de Luciano Floridi y J.W. Sanders, quienes no sólo argumentan que los agentes autónomos artificiales pueden considerarse agentes morales, sino también que constituyen un ejemplo de "moralidad sin mente" (2004, 351). En desacuerdo con Floridi y Sanders, sostengo que la sensibilidad es necesaria

Poole y Mackworth contribuye con una definición de IA muy *ad hoc* a este tema. Nos dice que la IA es el campo que estudia la síntesis y el análisis de agentes computacionales que actúan de manera inteligente. Y complementa: Un agente es algo (o alguien) que actúa. Un agente es inteligente cuando: 1. sus acciones son apropiadas para sus circunstancias y sus objetivos; 2. es flexible a entornos y objetivos cambiantes; 3. aprende de la experiencia, y 4. toma las decisiones adecuadas dadas sus limitaciones perceptivas y computacionales. También hay posturas que hacen la diferenciación entre agente inteligente (software) y el robot.

Por lo general, un agente artificialmente inteligente es un software que opera en línea o en un mundo simulado, a menudo generando percepciones y / o actuando dentro de este mundo artificial. Un robot, por otro lado, está situado en el mundo real, lo que significa que su existencia y operación ocurren en el mundo real. Los robots también están incorporados, lo que significa que tienen un cuerpo físico. El proceso de un robot que toma decisiones inteligentes a menudo se describe como sentido-plan-acto, lo que significa que el robot debe primero sentir el entorno, planificar qué hacer y luego actuar en el mundo.¹⁴¹

Sea uno u otro, el hecho es que los agentes computacionales o los agentes artificiales inteligentes al momento que toman decisiones, aunque no se compruebe aún la conciencia, intencionalidad y libre albedrío, deben ser considerados sujetos de responsabilidad algorítmica, al modificar por sí mismos el mundo o su entorno.

Anibal Monasterio Astobiza, nos dice la concepción estándar de agencia moral: “Esta concepción estándar de la agencia moral establece una conexión entre la acción y tener razones (obligaciones o requerimientos y capacidad para razonar sobre esos requerimientos). Así, de esta concepción se desprende el siguiente silogismo para definir la agencia moral de un agente: Si X desea A, entonces la acción B le lleva a conseguir A y por tanto X realizará B. Es evidente que además del problema de autonomía (autogobierno) y si el razonamiento es un proceso determinado o libre (distinción entre causas naturales y razones internas), existe el

para el albedrío moral, porque la experiencia consciente del tipo que permite sentimientos como el placer, el dolor y la empatía parece necesaria para el albedrío moral.

¹⁴¹ Cfr. Christoph Bartneck, Christophütge Alan Wagner Sean Welsh, *An Introduction to Ethics in Robotics and AI*, Springer, 2021, p. 8 y 12.

problema de identificar quién es un agente o no. La llegada de sistemas artificiales agudiza este problema y por supuesto en el caso de que llegáramos a construir máquinas autónomas con capacidad Ética: ¿cuáles serían nuestras obligaciones o deberes para con ellas?, ¿deberíamos respetar su autonomía?, ¿concederles derechos?” Un agente moral: 1) tiene obligaciones o requerimientos morales; 2) tiene libertad (autonomía); 3) puede razonar sobre esos requerimientos.¹⁴² Anibal responde: *No hay ninguna razón a priori para que no se pueda crear moralidad artificial*. Sin embargo... no hay ninguna línea de trabajo por parte de grupos de investigación... que intenten crear consciencia en máquinas o estados afectivos... no hay razón alguna que me haga pensar que la moralidad no sea computable, aunque sí que hay argumentos para pensar que la consciencia no es algorítmica.”.¹⁴³ Aníbal hace una tabla con las diferencias entre la Moral humana y la Moral artificial, que se inserta a continuación:

Diferencias entre Moral humana y Moral artificial

MORAL HUMANA	MORAL ARTIFICIAL
simpatía/empatía	?
procesos cognitivos de atención conjunta	?
procesos sociales/cognitivos de agencia y cooperación	x
procesos de auto-regulación	?
emociones	?
capacidad lingüística	?
cognición social	?
razonamiento causal	x
razonamiento contrafáctico	?
razonamiento de sentido común	x
autoconsciencia	x
ayuda a parientes	x
lealtad al grupo	x
reciprocidad	x
respeto a los superiores	x
ser valiente	x
dividir recursos en disputa	x
respetar la propiedad privada	x

Fuente: Anibal Monasterio Astobiza, *op. cit.*, p. 141

Por el momento -dice Carissa Véliz- los algoritmos se parecen a nosotros solo en algunas tareas y funciones que realizan en la sociedad, pero no tienen albedrío

¹⁴² Cfr. Aníbal Monasterio Astobiza, “Ética para máquinas: Similitudes y diferencias entre la moral artificial y la moral humana”, *ILEMATA, Revista Internacional de Éticas Aplicadas*, no. 30, 129-147, 2019, p. 132.

¹⁴³ *Ibidem.*, p. 143.

moral y en realidad no son autónomos como tal, “porque necesitan información externa para fijarse objetivos, ni respuestas a las razones, ya que ninguna razón podría “convencerlos” de cambiar el objetivo para el que han sido programados.”¹⁴⁴ Nosotros creemos que la comparación bajo la misma premisa no es aplicable. Estamos comparando cosas distintas (razonamiento hombre – razonamiento máquina). El hecho de que tomen decisiones por su cuenta y generen alguna modificación a las personas o su ambiente, los hace responsables o los haría responsables (civil, penal, administrativo, etcétera). No podemos clasificarlos o hacerlos responsables únicamente con el albedrío moral. Carissa Véliz menciona:

This paper has argued that moral zombies—creatures that behave like moral agents but lack sentience—are incoherent as moral agents. Only beings who can experience pain and pleasure can understand what it means to inflict pain or cause pleasure, and only those with this moral understanding can be moral agents. What I have dubbed ‘moral zombies’ are relevant because they are similar to algorithms in that they make moral decisions as human beings would—determining who gets which benefits and penalties—without having any concomitant sentience. There might come a time when AI becomes so sophisticated that robots might possess desires and values of their own. It will not, however, be on account of their computational prowess, but on account of their sentience, which may in turn require some kind of embodiment. At present, we are far from creating sentient algorithms. When algorithms cause moral havoc, as they often do, we must look to the human beings who designed, programmed, commissioned, implemented, and were supposed to supervise them to assign the appropriate blame. For all their complexity and flair, algorithms are nothing but tools, and moral agents are fully responsible for the tools they create and use.¹⁴⁵

¹⁴⁴ Carissa Véliz, “Moral zombies: why algorithms are not moral agents”, *AI & Soc*, 36, p. 435.

¹⁴⁵ *Ibidem.*, p. 496. “Este artículo ha argumentado que los zombis morales, criaturas que se comportan como agentes morales, pero carecen de sensibilidad, son incoherentes como agentes morales. Solo los seres que pueden experimentar dolor y placer pueden comprender lo que significa infligir dolor o causar placer, y solo aquellos con esta comprensión moral pueden ser agentes morales. Los que he llamado “zombis morales” son relevantes porque son similares a los algoritmos en el sentido de que toman decisiones morales como lo harían los seres humanos, determinando quién obtiene qué beneficios y sanciones, sin tener ninguna conciencia concomitante. Puede llegar un momento en que la IA se vuelva tan sofisticada que los robots puedan poseer deseos y valores propios. Sin embargo, no será por su destreza computacional, sino por su sensibilidad, que a su vez puede requerir algún tipo de encarnación. En la actualidad, estamos lejos de crear algoritmos sensibles. Cuando los algoritmos causan estragos morales, como sucede a menudo, debemos mirar a los seres humanos que los diseñaron, programaron, encargaron, implementaron y se suponía que los supervisarían para asignarles la culpa apropiada. A pesar de toda su complejidad y estilo, los algoritmos no son más que herramientas, y los agentes morales son totalmente responsables de las herramientas que crean y utilizan.”

Nosotros consideramos que el concepto de moralidad debe ser entendido como ese elemento que tienen los sistemas inteligentes o las máquinas para funcionar respetando las reglas Éticas que regulan la vida del hombre.

Un camino para regular la responsabilidad de los AAMA la encontramos en la creación de sistemas que supervisen en un segundo plano (o nivel superior) el comportamiento ético de los AAMA; es decir, crear una máquina que siga los principios de la Ética humana para su comportamiento y el de otras máquinas. Este esfuerzo ha sido planteado por un número de expertos que trabajan en el cómputo de teorías Éticas y/o del software que pone en práctica principios éticos en contextos particulares tales como cuidado médico o guerra. Algunos autores argumentan que las máquinas que son programadas con principios éticos cumplen más que los humanos.¹⁴⁶

Otro camino que se propone para regular la responsabilidad de los AAMA, ante la capacidad de aprendizaje de las Redes Neuronales Profundas y la posibilidad de que estos agentes no solo sean más capaces sino tengan mayores habilidades y menos posibilidades de interpretación humana, es hacerlos directamente responsables, ya que culpar a los programadores o usuarios, sin tener la posibilidad de demostrar la culpabilidad o no y el cómo surge la responsabilidad, podría llevar a muchas injusticias y al abuso del Derecho o al fraude a la ley, por personas que sepan esta imposibilidad y se dispongan a utilizarla para fines ilícitos.¹⁴⁷ En teoría, este tipo de AAMA, serán capaces de responder a elogios o a sentimientos, y por lo tanto podrán responsabilizarse de estas conductas. Johnston comenta "...no es que los robots serán responsables de su propio comportamiento, sino que los humanos se inclinarán a tratar a los robots como si los robots fueran responsables de su propio comportamiento" y, por tanto, se crearán mecanismos legales para el desahogo de las pruebas correspondientes.¹⁴⁸

Retomando a Johnson, coincidimos en considerar la posibilidad de que los robots sean responsables de su propio comportamiento si adquieren el estatus

¹⁴⁶ Johnson, D.G. Technology with No Human Responsibility?. J Bus Ethics 127, 2015, pp. 707–715. <https://doi.org/10.1007/s10551-014-2180-1>.

¹⁴⁷ *Idem*.

¹⁴⁸ *Idem*.

jurídico, no de personas sino, de AAMA, al menos en el sentido de que tendrán una especie de posición moral. Otros, continúa diciendo Johnson se manifiestan contra el uso abusivo de los robots y el respeto a sus derechos:

For example, Whitby (2008) argues that robots should not be abused and Petersen (2007) is concerned about the immorality of “robot servitude.” Although neither of these arguments suggests that robots of the future will or could have personhood, both arguments move in the direction of negative rights for robots. Moral standing does not necessarily mean responsibility for one’s behavior, but the arguments for moral standing could converge with the work being done to program artificial agents to be ethical. Establishing that artificial agents have moral standing and that they have the capacity to adhere to moral norms (i.e., to behave morally) would provide at least some of the groundwork for holding artificial agents responsible.¹⁴⁹

Sin duda, la responsabilidad jurídica es un concepto complejo, como lo vimos con Hart y Dworkin¹⁵⁰. Forma parte de un grupo de conceptos que tienen soporte moral. Dada la variedad de términos y sus significados, la responsabilidad se puede convertir en un punto de disputa entre aquellos que buscan cumplir la norma y aquellos que abusan de las omisiones de la norma. No entrar a este debate genera más afectaciones que beneficios a las nuevas tecnologías y a los AAMA. Al final del día, las grandes empresas productoras o consumidoras de AAMA están generando estrategias legales para mantener su crecimiento y productividad donde exista certeza jurídica. La no regulación de estos temas causa incertidumbre en el Derecho y aunque exista normas internas y de *soft law* que ya están aplicándose en muchas partes del mundo, la presencia del Estado es parte del ecosistema de protección social para el uso de la IA que se debe de construir (quizás varios, por regiones, por la propia complejidad humana).

¹⁴⁹ Cfr. Johnson, D.G. Technology with No Human Responsibility?, op. cit., pp. 707–715. <https://doi.org/10.1007/s10551-014-2180-1> “Por ejemplo, Whitby (2008) argumenta que los robots no deben ser abusados y Petersen (2007) está preocupado por la inmoralidad de la “servidumbre robot”. Aunque ninguno de estos argumentos sugiere que los robots del futuro tendrán o podrían tener personalidad, ambos argumentos se mueven en la dirección de los derechos negativos para los robots. La posición moral no significa necesariamente responsabilidad por el comportamiento de uno, pero los argumentos para la posición moral podrían converger con el trabajo que se está haciendo para programar agentes artificiales para que sean éticos. Establecer que los agentes artificiales tienen una posición moral y que tienen la capacidad de adherirse a las normas morales (es decir, comportarse moralmente) proporcionaría al menos parte del trabajo preliminar para responsabilizar a los agentes artificiales.”

¹⁵⁰ Ver páginas 382 a 392.

Lo que nadie duda es que los futuros robots tendrán habilidades superiores a los seres humanos, tanto en sus características físicas como en sus habilidades digitales. “En *Robot Futures*, la robótica Illah Reza Nourbakhsh considera cómo compartiremos nuestro mundo con esos seres artificiales, y cómo nuestra sociedad podría cambiar en la medida que incorporan seres más fuertes e inteligentes.”¹⁵¹ ¿Cómo el ser humano actuaría al ver que existe otro ente más inteligente y fuerte? ¿se convertirá en un animal más prudente, cuidadoso, cauto o respetuoso con el ser humano?

Creemos que los abogados debemos ir pensando sobre el tipo de responsabilidad que llegará y el cómo hacer que esta tecnología rinda cuentas sin afectar la innovación y el crecimiento del sector tecnológico ya que es indudable que éste aporta grandes beneficios al ser humano. Si puede haber tecnologías de las que ningún ser humano sea responsable, entonces qué mecanismo legal se debe utilizar para proteger sus Derechos Humanos de afectaciones por el uso de los sistemas inteligentes que deciden la participación del usuario, del programador o de la empresa. Cuando una tecnología se comporta de manera tal que sale del control humano lo que esperábamos, a menudo, es saber por qué ocurrió eso y quién es el responsable. Johnson concluye lo siguiente:

Artificial agents can be designed so that no human can understand or control what they do or they can be designed so that they are transparent and well within human control or they can be designed so that certain aspects or levels of the machine behavior are in human control and others are not. Which way they are designed depends on the humans involved in their development and acceptance. In the past people have chosen technologies that have some degree of risk though we have also set up mechanisms to pressure those who make and use these technologies to operate them safely and to take responsibility when something goes wrong and the fault can be traced back to them. The future may be different, but it seems there are good reasons why we might resist any future in which no humans are responsible for technologies that have a powerful role in our lives.¹⁵²

¹⁵¹ *Idem.*

¹⁵² *Idem.* “Los agentes artificiales pueden ser diseñados para que ningún humano pueda entender o controlar lo que hacen o pueden ser diseñados para que sean transparentes y estén bien dentro del control humano o pueden ser diseñados para que ciertos aspectos o niveles del comportamiento de la máquina estén en control humano y otros no. La forma en que están diseñados depende de los seres humanos involucrados en su desarrollo y aceptación. En el pasado, las personas han elegido tecnologías que tienen cierto grado de riesgo, aunque también hemos establecido mecanismos para presionar a quienes fabrican y utilizan estas tecnologías para que las operen de forma segura y asuman la responsabilidad cuando algo sale mal y la culpa se puede atribuir a ellas. El futuro puede ser diferente, pero parece que hay buenas razones por las que podríamos resistirnos a cualquier

Nuestra sugerencia es que el concepto AAMA puede llegar al Derecho, ya sea directa o indirectamente, así como el concepto responsabilidad algorítmica. Los abogados tenemos que acercarnos más a los filósofos (o eticistas) para entender estos temas y llevarlos a normas jurídicas. Los abogados también debemos aprender de los ingenieros y tomar sus consejos o conocimientos para sumarlos con los del juez, el fiscal o el perito, para la creación, enseñanza y aplicación del Derecho y la IA.

En resumen, para efectos de este trabajo entenderemos por agencia. - la capacidad de un agente (personal o de otra clase) para actuar en el mundo; agente moral. - cualquier persona capaz de actuar en el mundo; y, Agente Artificial Moral Autónomo (AAMA). - cualquier sistema o máquina capaz de actuar en el mundo sin intervención directa o indirecta del ser humano¹⁵³. Es una creación de la Ética para la IA. Este concepto permite dividir la responsabilidad que le corresponde a la máquina (responsabilidad algorítmica) de la responsabilidad que les corresponde a los seres humanos (responsabilidad objetiva, subjetiva, contractual, extracontractual, etcétera).

Esta nueva concepción del Agente y de la AAMA, permite construir y dar soporte a la categoría de responsabilidad algorítmica. En este tipo de tecnología, la máquina de manera independiente, con los datos que se le proporciona o que obtiene del ambiente, va proponiendo soluciones a temas específicos. Es decir, no existe la voluntad humana presente en la construcción de soluciones. La teoría de la autonomía de la voluntad y el libre albedrío, la capacidad de razonar, conocer y

futuro en el que ningún ser humano sea responsable de las tecnologías que tienen un papel poderoso en nuestras vidas.”

¹⁵³ Las definiciones se construyeron con base en: (1) Joaquín M. Fuster, *Cerebro y Libertad. Los circuitos cerebrales de nuestra capacidad para elegir*, México, Ediciones Culturales Paidós, 2019, p. 349 y ss. (2) JA., Cervantes, S., , López, L.F., Rodríguez, et al. “Artificial Moral Agents: A Survey of the Current Status”, *Sci Eng Ethics*, número 26, 2000, pp. 501-532. <https://doi.org/10.1007/s11948-019-00151-x> (3) C. Véliz, “Moral zombies: why algorithms are not moral agents”, *AI & Soc*, 36, 2021, 487-497. <https://doi.org/10.1007/s00146-021-01189-x> y (4) D.G. Johnson, “Technology with No Human Responsibility?” *J Bus Ethics*, 127, 2015, 707-715. <https://doi.org/10.1007/s10551-014-2180-1> (5) D., Howard, y I. Muntean, “A minimalist model of the artificial autonomous moral agent (AAMA)”, en *Symposium series AAAI Spring*, 2016, pp. 217-225. <https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/viewFile/12760/11954> (6) Consultese también el sitio: [Association for the Advancement of Artificial Intelligence \(aaai.org\)](https://www.aaai.org/) (20 de julio de 2022)

efectuar el acto con consciencia e inteligencia, a través de la voluntad del hombre, parecieran ser poco útiles para calcular la responsabilidad algorítmica. No hay un ser humano que consciente y voluntariamente haga una programación de la máquina que emite el resultado, ni un ser humano que esté guiándola en la búsqueda de información o el filtrado de datos. En el mejor de los casos, quizás encontremos a una persona que ingresa datos -el médico, su auxiliar o algún empleado que administra los equipos del hospital o que busca datos para que sean incorporados o incluidos a la base de datos o de información y éstos a su vez son analizados y procesados conforme la base de datos o de la información, el campo de la medicina o de la profesión que se tenga autorizado para su uso. Pero no existe aún, en México y en América Latina normas, instituciones y personas -físicas o jurídicas- que supervisen y auditen el uso de estos datos y de los algoritmos de inteligencia artificial que se utilizan para la toma de decisiones. Para determinar la responsabilidad algorítmica de un AAMA, se debe evaluar el tipo de algoritmo que se están utilizando, los datos, el modelo de IA, el tipo de algoritmo, el modelo de IA, si es opaco o no, si hubo o no sesgos algorítmicos, la interpretación y explicabilidad, entre otros.

Segundo concepto. - Responsabilidad algorítmica. Una vez hecha la definición de lo que es un AAMA ahora explicaremos lo que es la responsabilidad algorítmica. ¿Qué debemos entender por responsabilidad algorítmica? Ya que entendimos la necesidad de crear el concepto AAMA, que equivale al agente moral de la Filosofía occidental, el segundo concepto novedoso es el de responsabilidad algorítmica, el cual utiliza al primero concepto como su fuente o su origen y también al tipo de IA que utiliza. La responsabilidad algorítmica es un tipo de responsabilidad que no se tenía contemplada ni por los juristas del siglo XIX ni por los juristas del siglo XX, o por lo menos no con el panorama que estamos viviendo a partir del descubrimiento de sistemas inteligentes que generan soluciones automatizadas sin la intervención de la voluntad y el conocimiento humano.

Por lo menos en el mundo occidental, de 2016 a la fecha se ha venido dando un movimiento de muchos científicos, en su mayoría formados en Filosofía y Ética,

que ha cobrado gran relevancia para el Derecho y la forma en la que se debe regular la reparación del daño por una afectación a Derechos Humanos cuyo origen la encontramos en modelos automatizados para la toma de decisiones de los sistemas inteligentes basados en algoritmos como las Redes Neuronales Profundas.

Definimos a la responsabilidad algorítmica como aquella que se produce por un Agente Autónomo Moral Artificial (AAMA), que toma decisiones a través de sus propios algoritmos, sin la participación de la voluntad o decisión de un ser humano. El Parlamento Europeo, el 16 de febrero de 2017, creó la categoría jurídica de persona electrónica. Cuando la máquina toma decisiones de manera automática o autónoma con el solo procesamiento de los datos que obtiene de las bases de datos o del entorno, estamos ante una máquina que puede ser sujeto de responsabilidad; cuando la máquina no es capaz de aprender por sí sola y requiere de un ser humano, de su voluntad y consentimiento, estamos ante una categoría ajena a la responsabilidad algorítmica. La responsabilidad algorítmica es un concepto estudiado por la justicia algorítmica¹⁵⁴.

En la responsabilidad contractual tradicional, por ejemplo, hay un ser humano que da origen al acto de molestia que la ley sanciona o prohíbe. Existe, de manera natural y directa la manifestación de la voluntad y su libre albedrío. Puede o no puede haber una máquina que se utilice para cumplir la actividad o el servicio solicitado. Existe un ser humano afectado. Hay normas, instituciones, servidores públicos, abogados, defensores, peritos, fiscales y toda una estructura gubernamental para su atención, así como reglas procesales y conceptos que guían el actuar de cada una de estas personas, delimitándolas y supervisándolas para proteger la libertad contractual, la voluntad de las partes, el respeto a los Derechos Humanos, entre otros.

En cambio, en la responsabilidad algorítmica, que deriva de sistemas inteligentes para la toma de decisiones, basada en ciertos tipos de algoritmos, como, por ejemplo, las Redes Neuronales Profundas; no hay un ser humano que consciente y voluntariamente haga una programación de la máquina que emite el resultado, ni un ser humano que esté guiándola en la búsqueda de información o el

¹⁵⁴ Vid., p. 148.

filtrado de datos. En el mejor de los casos, quizás encontremos a una persona que ingresa datos -el médico, su auxiliar o algún empleado que administra los equipos del hospital o que busca datos para que sean incorporados o incluidos a la base de datos o de información y éstos a su vez son analizados y procesados conforme la base de datos o de la información el campo de la medicina o de la profesión que se tenga autorizado para su uso. Pero no existe aún, en México y en América Latina normas, instituciones y personas -físicas o jurídicas- que supervisen y auditen el uso de estos datos y de los algoritmos de inteligencia artificial que se utilizan para la toma de decisiones.

Se pudiera utilizar la Teoría de la Responsabilidad Jurídica expuesta en apartados anteriores y las reglas de nuestro sistema de justicia que veremos a continuación, pero el problema está en cómo comprobar el nexo causal, entre el hecho ilícito y la afectación que deriva de ese actuar prohibido. En otros apartados de este texto, se han mencionado propuestas para actualizar la legislación nacional y supranacional en diversas partes del mundo, que están buscando adicionar a la teoría tradicional de la responsabilidad penal, los ingredientes de la operación y el funcionamiento de sistemas inteligentes para la toma de decisiones, las cuales se están discutiendo. Pero lo ideal sería definir en norma la responsabilidad algorítmica o, *a contrario sensu*, definir en norma la obligación que tiene todo aquel ser humano de conocer previamente los riesgos que tiene el uso de un sistema inteligente.

La responsabilidad algorítmica también nos lleva a reflexionar sobre cómo sería la carga de la prueba, a quién le corresponderá hacerla y bajo qué formalidades. El incluir el concepto de responsabilidad algorítmica va de la mano con analizar el sistema vigente de la carga de la prueba y los antiguos estándares de prueba que se han venido utilizando y, relativamente, perfeccionando desde que Jeremy Bentham publicó su *Tratado de las pruebas judiciales* (1823)¹⁵⁵ cuando nos dice que hay una conexión que existe entre la ley y el modo de enjuiciar y del modo de enjuiciar con las pruebas. La obra final del juez es una *decisión*... cuando un

¹⁵⁵ Cfr. *Tratado de las pruebas judiciales*. Obra extraída de los manuscritos de M. Jeremías Bentham, *jurisconsulta inglés, escrita en francés*, por Estevan DUMONT, Vocal del Consejo representativo y soberano de Ginebra. Trad. C. M. V. Tomo I y II, Paris, Bossange Frères, 1825. [La obra original es de 1823, Paris] pp. 1-4.

particular recurre a un juez, es para pedirle una decisión... el deber del juez es el obtener las pruebas de una y otra parte, en la mejor forma posible, compararlas, y decidir... así pues, el arte de enjuiciar no es en substancia sino el arte de producir las pruebas. Por prueba se entiende un hecho supuesto verdadero, que se considera como que debe servir de motivo de credulidad sobre la existencia o no existencia de otro hecho... Toda decisión fundada en una prueba procede, pues, por vía de conclusión. Dado tal echo, concluyo que existe tal otro. En todos los casos la prueba es un medio para un fin. ¹⁵⁶

El sistema de valoración de la prueba se fue analizando y construyendo durante los siglos XIX y XX hasta llegar a la categorización que hoy conocemos. Por un lado, los estándares de prueba y por el otro la carga de la prueba (o sistemas de valoración de la prueba y carga de prueba). Tanto uno como el otro, pueden generar problemas al momento de la recopilación de indicios (evidencias) y su posterior evolución procesal (datos de prueba – pruebas) para la valoración de un juzgador. Los sistemas de valoración de la prueba, en el mundo occidental, utilizan el concepto anglosajón de estándar probatorio para marcar la pauta procesal del juez en esta materia y, a su vez, los estándares probatorios utilizan a la carga de la prueba para regular determinados actos procesales en los que las partes deben acreditar su dicho, pretensión o acción punitiva. ¹⁵⁷

¿Cómo sería la valoración de la prueba para demostrar la responsabilidad de un AAMA? Si partimos de que actualmente la valoración de la prueba es aquel acto mediante el cual se determina el valor probatorio de cada medio de prueba y su relación con cada hecho controvertido con el fin de ubicar circunstancias de modo, tiempo, lugar (cómo, cuándo, dónde, porqué, para qué, etcétera) y su relación con

¹⁵⁶ *Ibidem.*, pp. 19-21

¹⁵⁷ *Ibidem.*, p. 27. Bentham fija las bases de estudio de la prueba en materia civil y en materia penal. Entiende con mucha claridad la relación entre acto volitivo a través de la acción producida por el ser humano, que trae determinadas consecuencias que la ley prohíbe (o permite) y con ello, se llega al concepto de “culpabilidad” (hoy responsabilidad) de las personas. Dice Bentham: “El hecho de la culpabilidad de un hombre con relación a tal o cual delito, es una especie de hecho muy complejo. Se puede dividir desde luego en los hechos siguientes:

1º. Que el hombre ha cometido tal acción; 2º. Que tal acción estaba acompañada de tales o cuales circunstancias;

3º. Que había una ley que prohibía aquella acción en aquellas circunstancias; 4º. Que no había ley que permitiese aquella acción en ninguna de las circunstancias que ha sido cometida.”

el supuesto normativo. Es la fase en la que se va decidiendo un juicio ya que, una vez hecha esta valoración de la prueba, el juzgador hará su pronunciamiento sobre la litis que se le presentó para su resolución.

Históricamente la actividad del juzgado se ha venido regulando -o delimitando- en la legislación procesal de cada país que toman como base propuestas teóricas como la de prueba legal o tasada, prueba libre y lógica o modelos mixtos en los que se conjugan ambas (como el mexicano que en materia civil mantiene un sistema mixto¹⁵⁸ y en materia penal utiliza un sistema basado en la prueba libre y lógica¹⁵⁹).

La problemática que encontramos en ambos casos es que dichos sistemas fueron contruidos y pensados para la búsqueda de evidencias y la generación de pruebas en las que el ser humano tuvo una participación directa o a través de otro ser humano. Además, el estudio de la mecánica de aportación probatoria, tanto en materia civil como en el ámbito penal, en la mayoría de las veces no utiliza las aportaciones de la ciencia. Por eso varias voces académicas han manifestado su preocupación y están trabajando diversas propuestas para que los legisladores y juzgadores tomen más en cuenta las aportaciones científicas y las lleven al ámbito probatorio.¹⁶⁰

Aquí hay varios elementos que ya actualmente están siendo cuestionados, sin ser vinculados a investigaciones de hechos que se deben de probar para

¹⁵⁸ “PRUEBAS. EL OBJETIVO DEL SISTEMA DE VALORACIÓN LIBRE ES EL ESCLARECIMIENTO DE LOS HECHOS SIN NECESARIAMENTE BUSCAR LA VERDAD ABSOLUTA, SINO LA PROBABILIDAD MÁS RAZONABLE.” [TA]; 10a. Época; T.C.C.; Gaceta S.J.F.; Libro 77, agosto de 2020; Tomo VI; Pág. 6214 I.4o.A.44 K (10a.) Registro digital: 2021913

¹⁵⁹ “PRUEBAS EN EL SISTEMA PROCESAL PENAL ACUSATORIO. SU VALORACIÓN LIBRE Y LÓGICA POR EL JUZGADOR EN TÉRMINOS DEL ARTÍCULO 20, APARTADO A, FRACCIÓN II, DE LA CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 69, agosto de 2019; Tomo II; Pág. 1320 1a. LXXIV/2019 (10a.) Registro digital: 2020480

¹⁶⁰ Para conocer más sobre los textos académicos que han venido estudiados estos temas consúltense a: Erich Döhring, *La prueba, su práctica y apreciación*, trad. Banzhaf, Buenos Aires 1964. Hernando Evis Echandia, *Teoría General de la Prueba Judicial*, Tomos 1 a 6, Víctor P. de Zavallía Editor, Buenos Aires. (diversas fechas). Lluís Muñoz Sabaté, *Tratado de probática judicial*, Barcelona 1992. Juan Carlos, Cabañas García, *La valoración de las pruebas y su control en el proceso civil. Estudio dogmático y jurisprudencial*, Trivium, Madrid 1992. Michel Taruffo, *La prueba de los hechos*, trad. Ferrer, Marcial Pons, Madrid 2002. Marina Gascón Abellán, *Los hechos en el derecho. Bases argumentales de la prueba*, Marcial Pons, Madrid-Barcelona, 2004. Jordi Ferrer Beltrán, *La valoración racional de la prueba*, Marcial Pons, Madrid, 2007. Michele Taruffo, *La prueba*, Marcial Pons, Madrid, 2008. Carmen Vázquez (editora), *Estándares de prueba y prueba científica. Ensayos de epistemología jurídica*, Marcial Pons, Madrid, 2013. Larry Laudan, *Verdad, error y proceso penal*, Marcial Pons, Madrid, 2013.

demostrar la responsabilidad -civil o penal- de un sistema inteligente para la toma de decisiones en las que no participa un ser humano y que tiene que ver con diversos supuestos que se han venido presentando en la práctica cotidiana. En particular:

1. Cuando el juez interviene, en un sistema de valoración legal, no puede incluir hechos que el Derecho no le permite, solo debe creer lo que la ley dice que debe de cuidar para efectos del principio de contradicción procesal; o, por el contrario, si el juez interviene en el sistema es de valoración “libre y lógica, el juez utiliza su razonamiento, su leal saber y entender, las reglas de la sana crítica, las máximas de la experiencia, la íntima convicción, el grado de prueba, los estándares de prueba, etc.”.¹⁶¹
2. En los sistemas continentales, por lo general, la carga de la prueba parte del principio que los intervinientes deben de probar sus dichos y existen excepciones a esta regla que dependen de la materia, pero no hay una que obligue al creador del sistema de inteligencia artificial a que demuestre que no hubo afectación a Derechos Humanos ex ante o ex post. Un juez amante de la aplicación literal de la ley pudiera pedirle a la persona que demanda o denuncia una violación de sus Derechos Humanos, que muestre las evidencias de su dicho siguiendo las reglas que contiene el Código Civil, sin tener en cuenta que lo anterior es imposible hacerlo ya que no cuenta con la tecnología ni con los conocimientos para explicar la trazabilidad de estas evidencias, afectaciones o consecuencias en su esfera jurídica, dignidad, etcétera.¹⁶²
3. La carga de la prueba y el estándar de prueba atan de manos a la prueba científica, o por lo menos no están pensados para el desahogo de pruebas en cuestiones litigiosas que derivan del uso de sistemas inteligentes en las

¹⁶¹ Jordi Nieva-Fenoll, “Los sistemas de valoración de la prueba y la carga de la prueba: Nociones que precisan revisión”, en Justicia: revista de derecho procesal, No. 3-4, J. M. Bosch Editor, 2011, pp. 91-120.

¹⁶² Nieva-Fenoll comenta: “No existe siempre y en todos los casos una dicotomía entre prueba libre y prueba legal. En ocasiones, las normas de prueba legal sólo orientan la libre valoración, por lo que no son incompatibles con la misma. De hecho, debe recordarse que todas las actuales normas de prueba legal fueron un día simples orientaciones jurisprudenciales.” *Ibidem.*, p. 120.

que no participan de forma directa los seres humanos ya que una de las partes siempre queda con menos probabilidades de éxito al no poder probar su dicho. La llamada carga subjetiva de la prueba, también denominada carga formal, supone que cada parte debe aportar al proceso la prueba de los hechos que le son favorables. Y aún, en aquellos sistemas en donde el juez puede decidir de oficio, o a petición de parte, la distribución de la carga probatoria entre las partes, exigiendo a la parte que se encuentre en una situación más favorable para aportar las evidencias o esclarecer los hechos controvertidos (carga dinámica), se mantiene el problema de la trazabilidad de estas evidencias, afectaciones o consecuencias en su esfera jurídica, dignidad, etcétera.¹⁶³

Desde el Derecho Romano se pedía que cada una de las personas que intervenían en un juicio aportaran sus pruebas de las afirmaciones hechas (*onus probandi*).¹⁶⁴ Posteriormente, se inventan reglas sobre la inversión de la carga de la prueba, por ejemplo, la carga de la prueba dinámica, utilizada para determinados temas o materias y que complementan al *onus probandi*. Después llegaría la llamada carga objetiva o carga material, concepto que designa “no al litigante que debe probar un hecho sino directamente a los hechos que deben ser probados en el proceso para emitir un juicio... [y busca determinar] a qué parte perjudicará la falta de prueba de un hecho.”¹⁶⁵

¹⁶³ Jordi Nieva-Fenoll, “Carga de la prueba y estándares de prueba: dos reminiscencias del pasado”, en *InDret* 3. 2020, pp. 406-437. En su texto, Nieva-Fenoll cita al artículo 214 del Anteproyecto de Código Procesal Civil y Comercial de la Nación Argentina de 2019: “Artículo 214.-Carga de la prueba. Incumbirá la carga de la prueba a la parte que afirme la existencia de un hecho controvertido o de un precepto jurídico que el juez o el tribunal no tenga el deber. Incumbirá a cada una de las partes la carga de probar el presupuesto fáctico de las normas favorables a su pretensión, defensa o excepción. No obstante, según las particularidades del caso, el juez podrá, de oficio o a petición de parte, distribuir la carga al ordenar las pruebas, exigiendo probar determinado hecho a la parte que se encuentre en una situación más favorable para aportar las evidencias o esclarecer los hechos controvertidos. Se considerará en mejor posición para probar a la parte en virtud de su cercanía con el material probatorio, por tener en su poder el objeto de prueba, por circunstancias técnicas especiales, por haber intervenido directamente en los hechos que dieron lugar al litigio, o por estado de indefensión o de incapacidad en la cual se encuentre la contraparte, entre otras circunstancias similares. A la parte a quien se imponga la carga se le otorgará el plazo suficiente para ofrecer la prueba.”

¹⁶⁴ *Ibidem.*, 410.

¹⁶⁵ *Ibidem.*, 412.

Al mismo tiempo van surgiendo variedades de estándares probatorios. Los primeros en surgir fueron aquellos denominados de prueba legal, que funcionaba bajo 2 categorías principales:

- a) *Plena probatio* (estándar de convicción máxima que se podía alcanzar en el proceso).- Para llegar a éste se requería un documento público o dos testigos presenciales de los hechos; o bien a través del juramento. Jordi Nieva-Fenoll comenta que “detrás de un documento público había un fedatario que juraba que era cierto, o bien dos testigos que juraban o bien el propio litigante que juraba y esto hacía plena prueba.”¹⁶⁶
- b) *Semiplena probatio* (esta era una prueba indirecta que requería de una combinación de factores).- Las pruebas que no eran documentos públicos o que se ofrecían sin las reglas anteriores, se consideraban *semiplena probatio*. Y funcionaba de la siguiente manera -dice Nieva-Fenoll- un documento privado, un solo testigo o un litigante que firmaba sin juramento. Se sumaban el número de pruebas semiplenas para generar convicción en el juzgador.¹⁶⁷
- c) *Beyond any reasonable doubt* (más allá de toda duda razonable). - Tiene su origen en el Derecho Canónico y en la certeza moral”y lo encontramos en el *Common Law* (S. XVIII), concretamente para guiar el veredicto del jurado en juicios de orden criminal. Se mantiene en muchos países como el máximo estándar probatorio que se ha encontrado para generar seguridad jurídica al procesado.¹⁶⁸

Nieva-Fenoll, al estudiar el origen y evolución del estándar más allá de toda duda razonable comenta que la figura fue pensada no tanto como un estándar

¹⁶⁶ *Ibidem*. p. 418.

¹⁶⁷ *Idem*. Jordi Nieva-Fenoll hace una revisión exhaustiva de este tema citando, entre otros, a James Q. Whitman, *The origins of reasonable doubt*, New Haven, London, 2005.

¹⁶⁸ México lo incorporó a partir de la entrada en vigor del Código Nacional de Procedimientos Penales en el artículo 359: “Artículo 359. Valoración de la prueba. El Tribunal de enjuiciamiento valorará la prueba de manera libre y lógica, deberá hacer referencia en la motivación que realice, de todas las pruebas desahogadas, incluso de aquellas que se hayan desestimado, indicando las razones que se tuvieron para hacerlo. La motivación permitirá la expresión del razonamiento utilizado para alcanzar las conclusiones contenidas en la resolución jurisdiccional. Sólo se podrá condenar al acusado si se llega a la convicción de su culpabilidad más allá de duda razonable. En caso de duda razonable, el Tribunal de enjuiciamiento absolverá al imputado.”

probatorio, sino como una instrucción para jurados, "... se trataba de que expresaran su convicción según *the best of their Knowledge*, expresión inglesa corriente que está en la base de la concepción moderna del sistema de libre valoración. Pero conforme fue pasando el tiempo, el mundo del *Common Law* sintió... la necesidad de determinar mejor los niveles de convicción a fin de intentar evitar veredictos erróneos, y así surgieron la probabilidad preponderante - *preponderance of evidence*- o *balance of probabilities* en el Reino Unido y la *clear and convincing evidence*." ¹⁶⁹.

El policía, perito, fiscal o juez, cuando aplican cualquiera de estos estándares, con base en su norma jurídica, en realidad lo que buscan es cumplir un formalismo procesal que no busca nada científico, es un vestigio que traemos desde el momento en que el jurado se crea y éste, a la vez es considerado ya como una "reliquia histórica que por absurdas políticas se resiste a morir." ¹⁷⁰.

Las pruebas científicas, que se deben de utilizar para determinados actos basados en inteligencia artificial no deben de buscar la persuasión sin más, tampoco deben de buscar que el juez guíe su argumentación en cuestiones subjetivas, en vez de hacer un ejercicio de construcción basado en el Método Científico para que su validez tenga un soporte en la ciencia. Se requiere capacitar a todos los operadores jurídicos en estos temas y se necesita utilizar la propia inteligencia artificial para generar pruebas científicas y argumentos científicos. Si bien el juez no siempre sabrá de todos los avances tecnológicos, sí debe tener las bases y un equipo de apoyo multidisciplinario que de soporte a los poderes judiciales , y en su caso, colabore con el legislativo y el ejecutivo.

¹⁶⁹ Jordi Nieva-Fenoll., op. cit., p. 420. Jordi Nieva-Fenoll explica estas figuras de la siguiente manera: "La primera sería el estándar mínimo para dar por probado un hecho en una sentencia, que podría representarse intuitivamente con un porcentaje del 50,01% de convicción. El segundo sería más exigente, pero un escalón por debajo del "más allá de toda duda razonable". Surgen así, como se ve, tres estándares que recuerdan muy poderosamente a la *sempierna probatio* y a la *probatio plena*, quedando el *beyond any reasonable doubt* en el terreno de una especie de *probatio plenissima*, tan difícil de definir como aquel antiguo concepto. También se ha hablado de estándares inferiores a la *preponderance of evidence*, que sirven sobre todo en el proceso penal para iniciar actuaciones policiales. El más conocido sin duda es el de la *probable cause*, pero se han desarrollado algunos más sobre todo en la jurisprudencia estadounidense, pero que no son ni tan conocidos ni tan utilizados."

¹⁷⁰ Jordi, Nieva-Fenoll, op. cit., p. 422.

En síntesis, el sistema de valoración libre y lógica o el sistema de prueba tasada en norma, o la mezcla de ambos, requiere de añadir un eslabón para todo el tema de la recolección, ofrecimiento, depuración y admisión de pruebas científicas que estén relacionadas con el uso de sistemas inteligentes en los que la voluntad del ser humano no interviene para la toma de decisiones. Jordi Nieva-Fenoll lo entiende desde un enfoque evolutivo normativo y con un soporte filosófico:

“En todo caso, hemos llegado al imperio de la libre valoración de la prueba, lo que obliga a liberarse de todas las ataduras del pasado: estándares probatorios, carga de la prueba y una visión demasiado estricta del principio de aportación de parte. Queda para el futuro el resultado procesal que ello provocará, que se verá afectado además por otros factores relevantes, entre ellos el tecnológico. En todo caso, especialmente con este último habrá que estar muy pendiente de no perder la racionalidad a cambio de cálculos estadísticos sesgados y oscurantistas. El imperio de la razón fue una conquista social. Merecería la pena no perderlo.”¹⁷¹

La responsabilidad civil, penal, administrativa o cualquier otra que derive del sistema jurídico en el que se comenta el hecho ilícito o delito de los sistemas inteligentes que toman decisiones sin la intervención del ser humano debe pasar por un modelo probatorio que no encuentra su origen necesariamente en los medios probatorios tradicionales (confesional, testimonial, documental (pública o privada), pericial, presuncional o inspección judicial), ya que para hacer la trazabilidad del funcionamiento de los algoritmos hasta llegar a la decisión tomada (o solución aportada) requiere de otros sistemas de inteligencia artificial que intervengan en ese rastreo de evidencias y con los cuales las autoridades judiciales, fiscales, policías y peritos no cuentan. Su carencia ha sido solventada a través de la contratación de despachos especializados en la búsqueda de este tipo de evidencias.

Destaca también que actualmente no hay una sola persona que funja como perito especializado en elaboración de dictámenes de responsabilidad algorítmica. Ni ante el Consejo de la Judicatura Federal ni ante el Tribunal Superior de Justicia de la Ciudad de México, únicamente existen peritos en responsabilidad médica, en

¹⁷¹ Cfr. Jordi Nieva Fenoll, *Inteligencia artificial y proceso judicial*, Marcial Pons, Madrid, 2018, pp. 79 y ss. También consúltese, reflexiones complementarias sobre el uso de la carga de la prueba, carga dinámica de la prueba y principales problemas sobre su uso: Jordi Nieva Fenoll, Jordi Ferrer Beltrán y Leandro J. Giannini, *Contra la carga de la prueba*, Introducción de Michele Taruffo, Marcial Pons, Madrid, 2019.

medicina, en urgencias quirúrgicas, en medicina forense, en ciencias de la informática o en informática.¹⁷²

Tercer concepto, Medición de riesgos por el uso de IA.- Existe una tendencia mundial de regular y medir los riesgos de las organizaciones. Y esto ocurre también en el campo de la IA. El hecho que se vislumbra está en la inminente regulación de la IA basada en Machine Learning – Deep Learning. Si bien al principio había cierta resistencia y se pensaba que la regulación no era necesaria más allá de los programas de integridad corporativa o de *Compliance* para una autorregulación; a partir de 2018, diversos hechos cambiaron esta perspectiva y hoy las propias empresas exigen la intervención del gobierno y de sus facultades de control-legislación y control-supervisión. Los conceptos como opacidad, interpretabilidad, explicabilidad, sesgos algorítmicos, que no se encuentran en la legislación mexicana, ya están siendo analizados por los parlamentos y gobiernos. Estos también son conceptos que debemos medir en nuestro catálogo de indicadores.

No debemos olvidar que el 21 de abril de 2021, en el Parlamento Europeo, fue presentada la *Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*.¹⁷³ En su exposición de motivos se deja claro que si bien la mayoría de los sistemas de IA plantearán un riesgo bajo o nulo, algunos de esos sistemas entrañan peligros que deben regularse por los gobiernos para evitar situaciones no deseadas. Existen ya casos que han sido detectados que se pueden clasificar como nuevos riesgos relacionados con la seguridad de los usuarios y los derechos fundamentales. Esto afecta la seguridad jurídica y la posibilidad de que las personas y empresas adopten la IA de manera pronta debido a la falta de confianza, proponiendo cuatro niveles de clasificación basada en riesgos. Por su relevancia se transcribe a continuación:

“La Comisión propone un planteamiento basado en el riesgo, con cuatro niveles de riesgo: Riesgo inadmisibles: Se prohibirá un conjunto muy limitado de usos especialmente nocivos de la IA que contravienen los valores de la Unión al violar los derechos fundamentales (por ejemplo, puntuación social por parte de los Gobiernos,

¹⁷² https://www.iejcdmx.gob.mx/lista_peritos_auxiliares/

¹⁷³ Consúltase el texto en: https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683

explotación de los puntos débiles de los niños, uso de técnicas subliminales y, salvo contadas excepciones, determinados sistemas de identificación biométrica remota en directo en espacios públicos con fines policiales). Alto riesgo: Se considera de alto riesgo un número limitado de sistemas de IA definidos en la propuesta y que tienen un impacto negativo en la seguridad de las personas o en sus derechos fundamentales (protegidos por la Carta de los Derechos Fundamentales de la UE). Se adjunta a la propuesta la lista de sistemas de IA de alto riesgo, que puede revisarse para adaptarla a la evolución de los casos de uso de la IA (a prueba de futuro). Entre ellos figuran los componentes de seguridad de los productos contemplados en la legislación sectorial de la Unión. Siempre serán de alto riesgo cuando estén sujetos a una evaluación de la conformidad por terceros con arreglo a dicha legislación sectorial. A fin de garantizar la confianza y un nivel elevado y coherente de protección de la seguridad y los derechos fundamentales, se proponen requisitos obligatorios para todos los sistemas de IA de alto riesgo. Esos requisitos se refieren a la calidad de los conjuntos de datos utilizados; la documentación técnica y la llevanza de registro; la transparencia y la divulgación de información a los usuarios; la supervisión humana, y la solidez, la precisión y la ciberseguridad. En caso de infracción, los requisitos permitirán a las autoridades nacionales acceder a la información necesaria para investigar si el uso de la IA cumplió la legislación. La normativa propuesta es coherente con la Carta de los Derechos Fundamentales de la Unión Europea y está en consonancia con los compromisos internacionales de la Unión en materia de comercio. Riesgo limitado: En el caso de determinados sistemas de IA se imponen obligaciones específicas de transparencia, por ejemplo, cuando exista un riesgo claro de manipulación (por ejemplo, mediante el uso de robots conversacionales). Los usuarios deben ser conscientes de que están interactuando con una máquina. Riesgo mínimo: Todos los demás sistemas de IA pueden desarrollarse y utilizarse con arreglo a la legislación vigente sin obligaciones jurídicas adicionales. La inmensa mayoría de los sistemas de IA utilizados actualmente en la UE pertenecen a esta categoría. De forma voluntaria, los proveedores de estos sistemas pueden optar por aplicar los requisitos de una IA digna de confianza y adherirse a códigos de conducta voluntarios.”¹⁷⁴

¹⁷⁴ Para los efectos de nuestra investigación nos interesan los sistemas de alto riesgo ya que ahí se encuentra la IA basada en Machine Learning – Deep Learning enfocada a la salud. Una IA de riesgo alto incluyen la tecnología de IA utilizada en:

- Infraestructuras críticas (por ejemplo, transporte), que podrían poner en riesgo la vida y la salud de los ciudadanos;
- Formación educativa o vocacional, que puede determinar el acceso a la educación y el curso profesional de la vida de alguien (por ejemplo, calificación de exámenes);
- Componentes de seguridad de los productos (por ejemplo, aplicación de IA en cirugía asistida por robot);
- Empleo, gestión de trabajadores y acceso al trabajo por cuenta propia (por ejemplo, software de clasificación de CV para los procedimientos de contratación);
- Servicios públicos y privados esenciales (por ejemplo, calificación crediticia que niega a los ciudadanos la oportunidad de obtener un préstamo);
- Aplicación de la ley que pueda interferir con los derechos fundamentales de las personas (por ejemplo, evaluación de la confiabilidad de las pruebas);
- Gestión de migración, asilo y control de fronteras (por ejemplo, verificación de la autenticidad de los documentos de viaje);
- Administración de justicia y procesos democráticos (por ejemplo, aplicación de la ley a un conjunto concreto de hechos).

Los sistemas de IA de alto riesgo estarán sujetos a estrictas obligaciones antes de que puedan comercializarse:

- Sistemas adecuados de evaluación y mitigación de riesgos;
- Alta calidad de los conjuntos de datos que alimentan el sistema para minimizar los riesgos y los resultados discriminatorios;

Es muy importante tener presente la clasificación de riesgos que se está proponiendo en la Unión Europea y vincularla con los riesgos que afectan Derechos Humanos. La creación de legislación sin generar mecanismos de supervisión es, en muchos temas, letra muerta. Por ello, coincidimos con aquellos autores que están proponiendo la aplicación de auditorías algorítmicas de IA basada en Machine Learning – Deep Learning. Sin lugar a duda este tipo de acciones de supervisión ha traído grandes resultados a muchos sectores de la producción y los servicios. Sobre todo, en el ámbito financiero y de los servicios de salud. Partamos de una pregunta: ¿Cómo determinar en una auditoría algorítmica qué es bueno y qué es malo para el ser humano? Como hemos dicho, en esto ha habido varios caminos. El camino de los datos y la privacidad, el camino de la Ética de la IA y, un tercer camino está en los Derechos Humanos. Para llegar al primero analizaremos las propuestas de auditorías algorítmicas basadas en un enfoque de riesgos y terminaremos explicando el modelo para una Evaluación de Impacto en Derechos Humanos (EIDH).

Creemos también necesario reiterar que el 28 de febrero de 2020, en el Vaticano, se reunieron el presidente de IBM (John Kelly III), el presidente de Microsoft (Brad Smith), el director de la FAO (Qu Dongyu), el presidente de la Academia de la Vida (Vincenzo Paglia) y la Ministra Italiana para la Tecnología, Innovación y digitalización (Paola Pisano), teniendo como testigo al presidente del Parlamento Europeo (David Sassoli) para firmar el documento *Rome call for AI Ethics*.¹⁷⁵ La objetivo de la reunión fue mantener a la IA con un sentido humanista. Es decir, respetar la Declaración Universal de los Derechos Humanos, buscar el desarrollo de todas las personas, cuidar su dignidad y el medio ambiente, así como las necesidades de los más vulnerables.

-
- Registro de actividad para garantizar la trazabilidad de los resultados;
 - Documentación detallada que proporcione toda la información necesaria sobre el sistema y su propósito para que las autoridades evalúen su cumplimiento;
 - Información clara y adecuada al usuario;
 - Medidas apropiadas de supervisión humana para minimizar el riesgo;
 - Alto nivel de robustez, seguridad y precisión.

¹⁷⁵ Rome Call for AI Ethics se puede consultar en: https://www.romecall.org/wp-content/uploads/2021/02/AI-Rome-Call-x-firma_DEF_DEF_con-firme_.pdf

Los hechos que genere la IA deben ser el punto de referencia para una Ética digital y la promoción del uso de la tecnología en beneficio de a la humanidad y el medio ambiente. Todos los seres humanos nacen libres e iguales en dignidad y derechos. Están dotados de razón y conciencia, deben comportarse fraternalmente los unos con los otros (art.1 DDHH). Esta condición fundamental de libertad y dignidad -dice la declaración- debe ser también protegida y garantizada cuando se producen y usan los sistemas de IA.

Para ello deben existir tres requisitos: 1) Incluir a todo ser humano, sin ninguna discriminación; 2) debe hacer cosas buenas para la humanidad y para cada ser humano; y 3) debe ser consciente de la realidad compleja de nuestro ecosistema y estar definida por el modo en el cual cuida y protege el planeta (uso sustentable). La obligación de hacer sistemas transparentes/explicables es un criterio básico de la IA y de los algoritmos de agentes inteligentes. Los principios contenidos en esta declaración son:

- *Transparency* (transparencia): en principio, los sistemas de IA deben ser explicables.
- *Inclusion* (inclusión): La necesidad de que todos los seres humanos sean tomados en cuenta para que cada uno pueda beneficiarse y que a todos los individuos le sean ofrecidas las mejores condiciones para expresarse y desarrollarse.
- *Responsibility* (responsabilidad) aquellos que diseñan y despliegan el uso de la IA deben proceder con responsabilidad y transparencia.
- *Impartiality* (imparcialidad) no crear o actuar de acuerdo con sesgos, sino salvaguardar la justicia y la dignidad humana;
- *Reliability* (fiabilidad) los sistemas de IA deben ser capaces de trabajar fiablemente;
- *Security and privacy* (seguridad y privacidad): los sistemas de IA deben trabajar con seguridad y respeto por la privacidad de los usuarios.¹⁷⁶

¹⁷⁶ El texto original dice: “For this reason, the sponsors of the call express their desire to work together, in this context and at a national and international level, to promote “algor-ethics”, namely the ethical use of AI as defined by the following principles:1.Transparency: in principle, AI systems must be explainable;2.Inclusion: the needs of all human beings must be taken into consideration so

Algunos han manifestado que se debe poner un protocolo de gestión de disputas y un modelo del control y gestión de riesgos para mitigar las amenazas naturales de la tecnología y su uso en el mundo y, a la vez, la exigencia a las empresas que trabajan con IA de un Certificado de Integridad Ética para demostrar que están cumpliendo los estándares que cada país o región determine. Esto obliga a trabajar con modelos de medición como las matrices de riesgo y sobre programas de responsabilidad basada en riesgos (*risk-based liability*) o *Accountability – Due Diligence*.¹⁷⁷

Para concluir esta sección nos preguntamos ¿Sería justo hacer responsable a un ser humano por el comportamiento errado de una máquina que no controla? El llamado “requisito de control”¹⁷⁸ del sistema o la máquina que el ser humano utiliza para ciertos fines es clave para entender el tema. Según este concepto no es justo responsabilizar a alguien por los resultados o el comportamiento que no pudo controlar. Sin embargo, en la legislación vigente existen hipótesis normativas que te llevan a responsabilizar al ser humano por los resultados que no pudieron controlar.

La responsabilidad objetiva es un ejemplo de ello. Si se tomara la responsabilidad vigente y la aplicáramos a este tipo de casos, los ingenieros o las empresas para las que trabajan serían considerados responsables del comportamiento de los AAMA ya sea por motivos de responsabilidad profesional, ya sea por motivos de una relación de subordinación (patrón-trabajador). En ambos casos, se responsabiliza a personas (físicas o jurídicas) a pesar de que no son capaces de controlar el resultado. Esta ha sido la salida que tienen los sistemas jurídicos.

Una postura sobre este tema la expusimos cuando hablamos de Dworkin, quien dice que hay una interpretación alternativa del significado de tener el control. Según este punto de vista “un agente tiene el control cuando es consciente de

that everyone can benefit and all individuals can be offered the best possible conditions to express themselves and develop;3.Responsibility: those who design and deploy the use of AI must proceed with responsibility and transparency;4.Impartiality:do not create or act according to bias, thus safeguarding fairness and human dignity;5.Reliability: AI systems must be able to work reliably; 6.Security and privacy: AI systems must work securely and respect the privacy of users.”

¹⁷⁷ Ver apartado “Regulación algorítmica” de esta investigación.

¹⁷⁸ Cfr. Johnson, D.G. Technology with No Human Responsibility?. *J Bus Ethics* 127, 2015, 707–715. <https://doi.org/10.1007/s10551-014-2180-1>

enfrentar y tomar una decisión, cuando ningún otro toma esa decisión por su intermedio y para él y cuando el propio agente tiene las capacidades para forjarse creencias verdaderas sobre el mundo y hacer coincidir sus decisiones con su personalidad normativa: sus deseos, ambiciones y convicciones establecidas. Este es el sentido del control basado en la capacidad.”¹⁷⁹.

Siguiendo a Dworkin, podemos decir que existe el principio del control casual que ve la cuestión de la responsabilidad al margen de la percepción corriente que un agente tiene de su situación; y, el principio del control basado en la capacidad que pone la responsabilidad entre los paréntesis de una vida corriente vivida desde una perspectiva personal. Nosotros creemos, que el uso de estas nuevas tecnologías obliga a la sociedad y/o gobiernos a actualizar las normas vigentes y a realizar una ampliación con aquellas categorías que no existen actualmente en los sistemas jurídicos, por lo que los AAMA serían uno de esos conceptos necesarios que se deben estudiar para valorar esta ampliación. Si reconocemos que existen sistemas o máquinas capaces de actuar en el mundo sin intervención directa o indirecta del ser humano y lo vinculamos a nuestro Derecho vigente con alguna otra categoría, como por ejemplo la responsabilidad de la empresa o la responsabilidad por el uso de máquinas, se podría ir perfeccionando el andamiaje jurídico para lograr resolver casos difíciles que seguramente se presentarán en un futuro no muy lejano.

El Derecho Penal y la posibilidad de que los AAMA sean responsables ante sus reglas es un escenario muy tentativo. En el contexto del Derecho Penal surge la cuestión de la personalidad de los robots o de los sistemas que toman decisiones. Así como en su momento el Derecho Penal solo reconocía a la persona física como sujeto para la imposición de penas y hoy, adicionalmente, encontramos a las personas jurídicas (aclarando que no cambió la Teoría del Delito ni los elementos del tipo, sino que fue complementada con una categoría específica ad hoc a la naturaleza de las empresas), así ahora los sistemas inteligentes y robots (AAMA), serían considerados sujetos del Derecho Penal.

La principal razón de peso la encontramos en la propia justificación del cambio normativo en comento: Al igual que las personas jurídicas son utilizadas

¹⁷⁹ Para entender con mayor detalle el tema remítase a la sección “Dworkin” de este texto.

para cometer delitos, los robots pueden ser una entidad no humana que también cometa delitos y, por lo general, este tipo de tecnología la encontramos muy vinculada al quehacer empresarial (habría una Responsabilidad Penal de los AAMA por un daño al orden financiero, de salud, económico, político, educativo, laboral, entre otros; en una comunidad, país, región, continente o en todo el mundo) con consecuencias que van desde la suspensión del uso parcial o temporal del robot o sistema hasta la desconexión o destrucción permanente del mismo)¹⁸⁰

Así como el Derecho Penal evolucionó en el tema de la acción, la culpabilidad o la finalidad de la pena, creados en principio para la persona física, así debe evolucionar para regular el uso de sistemas inteligentes.¹⁸¹ Es una realidad, que la sociedad contemporánea usa a la empresa para delinquir y la empresa utiliza a los AAMA como herramienta para determinadas formas de criminalidad que se han detectado en el sector económico, financiero, de la seguridad nacional y en las políticas gubernamentales. La mayoría de los delitos contra el orden económico o financiero son cometidos con ayuda de IA; gracias a ella se mueven millones de dólares diariamente y se hacen operaciones comerciales y financieras que pasan desapercibidas en su licitud o ilicitud.

Con este tema concluiríamos los nuevos conceptos que tienen que analizarse para su incorporación a cualquier sistema jurídico. Ahora estudiaremos las categorías de la responsabilidad moral.

¹⁸⁰ Se puede definir delito económico como: “Aquella infracción que, afectando a un bien jurídico patrimonial individual, lesiona o pone en peligro, en segundo término, la regulación jurídica de la producción, distribución y consumo de bienes y servicios” Cfr. Miguel Bajo Fernández y Silvina Bacigalupo Saggese, *Curso de derecho penal económico*, Editorial Universitaria Ramón Areces, Madrid: 2010. Un libro novedoso que sirve de complemento para estos temas es: *Inteligencia artificial y Derecho: Desafíos y perspectivas*, Michelle Azuaje y Pablo Contreras (editores), Tirant lo Blanch, Valencia, 2021.

¹⁸¹ Para conocer más sobre esta evolución consúltese: Silvina Bacigalupo Saggese, *La responsabilidad penal de las personas jurídicas. Un estudio sobre el sujeto del Derecho Penal*, Universidad Autónoma de Madrid, 1997 [Tesis para la obtención del grado de Doctora en Derecho bajo la dirección del Dr. D. Gonzálo Rodríguez Mourullo] y Lina Mariola Díaz Cortés, “Societas delinquere Potest” Hacia un cambio de paradigma en el derecho penal económico, Aproximación al estudio de los delitos informáticos, *Derecho Penal Contemporáneo*, Revista Internacional, p. 79-114.

C) Categorías de la responsabilidad moral.

Hemos concluido el recorrido de argumentos relacionados con la responsabilidad jurídica y sus diferentes categorías (vigentes y no vigentes). Ahora entraremos a la revisión de la responsabilidad moral, sus características y el cómo se ha venido utilizando. La exposición de este apartado se hará en dos partes. La primera, con conceptos básicos de la Ética tradicional de Hans Jonas, por estar estos vinculados al uso de la tecnología; la segunda, con el pensamiento de la Ética empresarial de Carlos Llano Cifuentes.

Hans Jonas (1903-1993) fue discípulo de Edmund Husserl y de Martin Heidegger, amigo de Hannah Arendt. Carlos Llano Cifuentes (1932-2010), mexicano, de padre español y madre cubana, uno de los fundadores del IPADE y rector de la Universidad Panamericana, así como emprendedor de muchos otros proyectos académicos y empresariales.¹⁸²

Hans Jonas (1903-1993), lleva el tema de la responsabilidad al uso de la tecnología y los problemas éticos que genera. En su obra clásica *El principio de la responsabilidad* (1979) nos muestra un panorama sobre el ser humano y las amenazas de la tecnología creando una propuesta de imperativo categórico basado en el respeto de la libertad del hombre, la preservación de la integridad del medio ambiente y los límites del poder (*Tractatus technologico-ethicus*).

En su texto encontramos la inspiración aristotélica¹⁸³ y kantiana¹⁸⁴ y le da un enfoque muy vanguardista para la época de su publicación. Parte de dos preguntas:

¹⁸² Para conocer un panorama general de su vida, obra y aportaciones consúltese Héctor Zagal, *Para entender a Carlos Llano*, FCE, México, 2014. Existe también la Cátedra Carlos Llano UP-IPADE, como proyecto de divulgación de la obra de Llano en el campo de la investigación, la difusión y la docencia. En la Biblioteca del IPADE se puede consultar su obra sobre Management y sus textos filosóficos. <https://www.carlosllanocatedra.org/>

¹⁸³ Jonas comenta: "No todos los teóricos de la Ética han llevado tan lejos la reducción del aspecto cognoscitivo de la acción moral. Pero ni siquiera cuando se le ha otorgado gran importancia -como ocurre en Aristóteles, donde el conocimiento de la situación y lo que a ella conviene introduce considerables exigencias de experiencia y juicio- tiene ese saber nada que ve con la ciencia teórica. Naturalmente ese saber encierra en sí un concepto general del bien humano como tal, referido a las supuestas constantes de la naturaleza y condición humanas, y ese concepto general del bien puede o no ser elaborado en una teoría propia." Cfr. Hans Jonas, *op. cit.*, p. 31.

¹⁸⁴ Respecto a Kant, Jonas nos dice: "Kant fue tan lejos como para afirmar que "la razón humana puede llegar en lo moral, aun con el más vulgar entendimiento, a una gran exactitud y acierto"; que "no se precisa ciencia o Filosofía alguna para saber lo que se tiene que hacer, para ser bueno y honrado e incluso sabio y virtuoso... [El entendimiento vulgar puede] abrigar la esperanza de acertar, del mismo modo que un filósofo puede equivocarse"; "No necesito una gran agudeza para conocer

¿por qué debe haber hombres en el mundo? ¿por qué debemos garantizar su existencia futura?¹⁸⁵. y añade que la tecnología se ha convertido en una amenaza. “El hombre ha tenido un éxito desmesurado en sus investigaciones tecnológicas, que afecta a la propia naturaleza humana y lo coloca en una posición de gran responsabilidad. Los cambios tecnológicos son diferentes a lo ya experimentado antes. No hay punto de comparación. Ello hace que ninguna de las Éticas habidas hasta ahora nos instruya acerca de las reglas de bondad y maldad a las que las modalidades enteramente nuevas del poder y de sus posibles creaciones han de someterse. La tierra virgen de la praxis colectiva en que la alta tecnología nos ha introducido es todavía, para la teoría Ética, tierra de nadie.”.¹⁸⁶ Esto representa un vacío de valores que pone en peligro al ser humano y a la naturaleza y aquí es donde entra el concepto responsabilidad y la preocupación de Jonas:

La responsabilidad nunca antes tuvo un objeto de tal clase y hasta ahora había ocupado poco a la teoría Ética. Tanto el saber cómo el poder eran demasiado limitados como para incluir en su previsión el futuro remoto y para incluir en la conciencia de la propia causalidad el globo terráqueo... **la Ética se concentraba en la cualidad moral del acto momentáneo mismo, en el que había de tenerse en cuenta el derecho del prójimo que con nosotros convivía.** Mas, bajo el signo de la **tecnología, la Ética tiene que ver con acciones** -si bien ya no las del sujeto individual- **de un alcance casual que carece de precedentes y que afecta al futuro;** a ello se anda unas capacidades de predicción, incompletas como siempre, pero que superan todo lo anterior”.¹⁸⁷

Jonas también habla de la responsabilidad y la Ética en sus conceptos tradicionales. Por ejemplo, respecto de la Ética ubica no una sino varias Éticas que se han construido a lo largo del tiempo y que tienen premisas particulares: 1) Condición humana; 2) Bien humano; y, 3) Actuar humano con sus alcances. En conjunto se perfila hacia la responsabilidad del ser humano. El problema en estos ingredientes es que no son tan válidos para evaluar conductas en donde no participa el ser humano de forma directa y en las que no hay presencia física de la persona.

lo que tengo que hacer para que mi voluntad sea moralmente buena. Inexperto respecto al curso del mundo, incapaz de tomar en cuenta todo lo que en él acontece” puedo, sin, embargo saber cómo debo actuar conforme a la Ley moral”. Cfr. Hans Jonas, *op. cit.*, p. 30-31.

¹⁸⁵ Hans Jonas, *op. cit.*, p. 15.

¹⁸⁶ *Ibidem.*, p. 16.

¹⁸⁷ *Ibidem.* pp. 16-17.

Las nuevas o modificadas acciones humanas -dice Jonas- exigen un cambio de Ética. No es que la Ética tradicional y sus ingredientes deban de desaparecer, sino que debe ampliarse o evaluarse los ingredientes que permitirá aplicar las reglas de comportamiento válidas para el uso de las nuevas tecnologías. Los nuevos casos problemáticos que están surgiendo por el uso de Redes Neuronales Profundas son los que nos van a dar la pauta para analizar, proponer y evaluar los ingredientes de la Ética de la Inteligencia Artificial y su influencia en la responsabilidad de estas máquinas.¹⁸⁸

Debemos tener presente que las capacidades físicas e intelectuales del ser humano que está naciendo en el siglo XXI no son las mismas capacidades físicas e intelectuales que tenía el hombre de los siglos XVIII, XIX y XX. Los avances tecnológicos están influyendo en su estructura, desde antes que nazca y con el alargamiento de su vida a edades poco imaginadas en épocas pretéritas. Tan sólo pensemos en los distintos medicamentos que existen y están a la mano de muchas personas para ampliar sus capacidades de memoria, fuerza, percepción, intuición y que, en conjunto, hacen necesaria una revisión de conceptos como el de agencia moral y responsabilidad, para proteger los bienes jurídicos tutelados con los que el humano nace o aquellos que va adquiriendo a lo largo de su vida (respeto a su nacimiento, vida, salud, familia, desarrollo, construcción de su propio destino, etcétera)¹⁸⁹.

¿De qué forma afecta todo esto a la naturaleza humana y a la naturaleza del planeta? Si bien el ser humano siempre ha creado tecnología, lo que ahora influye

¹⁸⁸ En caso de aceptar, en algún momento, que son agencias o personas electrónicas que cambian el mundo con sus decisiones y que ese cambio, de ser violatorio de Derechos Humanos, debe ser considerado como un supuesto que encuadra en la responsabilidad algorítmica. Véase Cfr. Javier Cuairán, “La responsabilidad penal de los nuevos entes de inteligencia artificial”, *El país, sección Cinco días*, 1 de junio de 2021, versión digital. Consúltase en: https://cincodias.elpais.com/cincodias/2021/05/31/legal/1622463707_166390.html

¹⁸⁹ Para conocer datos sobre los avances en medicamentos que mejoran las capacidades de las personas, sus beneficios, repercusiones, la relación con conceptos tales como agente moral, libre albedrío, autonomía de la voluntad, consúltase la interesantísima investigación doctoral de María Isabel Cornejo Plaza: *Neurociencia y Derecho. El caso de la mejora cognitiva farmacológica*, Tesis de Doctorado, Universidad de Chile. Repositorio Institucional- Facultad de Derecho, Universidad de Chile, 2021. Respecto el tema del llamado “transhumanismo”, consúltase a: Antonio Diéguez, *Transhumanismo. La búsqueda tecnológica del mejoramiento humano*, Herder, Barcelona, 2017. Y como complemento: Manuel Calvo, *Filosofía para la era digital*, Almuzara, Madrid, 2018.

es esa potencialidad de uso a través de sistemas inteligentes que deciden sin la intervención humana directa. La solución va de la mano con el tipo de responsabilidad que ahora se debe categorizar.

Jonas también analiza un tema que no debemos perder de vista y que tiene que ver con lo permanente y lo cambiante. La Ética -dice nuestro autor- estaba en una posición permanente enfocada al *hic et nunc*, en la toma de conciencia y en la resolución de problemas de la realidad. Señalaba o se guiaba por circunstancias de modo, tiempo y lugar específicas que imputaban a su actor alguna responsabilidad. Sin embargo, muchos actos del hombre y de las máquinas que ha creado están alejándose de esta categoría y ese complemento se debe ir construyendo por esta generación para los que están por venir.

¿Qué características tenía la Ética tradicional o cuáles debe tener ahora? La Ética tradicional -a decir de Jonas- tiene cuatro características: 1. Todo trato con el mundo extrahumano era Éticamente neutro (a excepción de la medicina); tanto con relación al objeto como con relación al sujeto de la acción. Antes la técnica no afectaba la integridad del objeto, hoy lo hace de múltiples maneras; y con relación al sujeto, porque la *techne* (capacidad productiva, arte) y la modificación de los objetos no humanos no constituía un ámbito de relevancia Ética. 2. El trato directo entre seres humanos era lo que tenía relevancia Ética. 3. El hombre como entidad, no era visto como objeto de la *techne* (capacidad productiva, arte) transformadora. 4. El análisis de la conducta buena o mala no era un asunto de planificación o consecuencias lejanas. El lapso y las consecuencias de los actos eran determinados con claridad, tenían un inicio y un fin y una posible retribución de consecuencias por la responsabilidad del actor. Hay un entorno o relación inmediata.¹⁹⁰

¹⁹⁰ Jonas da ejemplos de máximas Éticas de tipo directo y con consecuencia inmediatas: “Ama a tu prójimo como a ti mismo”; “No hagas a los demás lo que no desees que te hagan a ti”; “Educa a tu hijo en el camino de la verdad”; “Busca la excelencia mediante el desarrollo y la realización de las mejores posibilidades de tu ser como hombre”; “Antepón el bien común a tu bien particular”; “No trates nunca a los hombres solamente como medios, sino siempre también como fines en sí mismos”... “en estas máximas el agente y “el otro” de su acción participan de un presente común. Quienes tienen algún derecho sobre mi comportamiento, en la medida en que mi acción u omisión los afecta, son los que ahora viven y tiene algún trato conmigo... Toda moralidad quedaba reducida a este estrecho campo de acción”. Hans Jonas, *op. cit.*, p. 30.

En el modelo tradicional de responsabilidad hay algo que se llama autoría de la acción inmediata directa: “a nadie se le hace responsable de los efectos posteriores no previstos de sus actos bienintencionados, bien-meditados y bien-ejecutados. El corto brazo del poder humano no exigía ningún largo brazo de un saber predictivo; la parvedad de uno era tan poco culpable como la del otro.”¹⁹¹

Sin embargo, dice Jonas, hay un momento en el que se pondrá al *homo faber* por encima del *homo sapiens* y es aquí donde la técnica moderna comienza a influir en la vida del hombre y en su responsabilidad. La tecnología “cobra significación Ética por el lugar central que ocupa ahora en la vida de los fines subjetivos del hombre. Poco a poco el mundo artificial va extendiéndose, el primer modelo lo encontramos en las ciudades, en donde las obras de la tecnología convierten al hombre y su mundo”¹⁹².

Jonas detecta bien lo que por décadas ha venido ocurriendo: “...el hombre es ahora cada vez más el producto de aquello que él ha producido, el hacedor de aquello que él puede hacer y, sobre todo, el preparador de aquello que en breve él será capaz de hacer... y es el futuro indeterminado más que el espacio contemporáneo de la acción el que nos proporciona el horizonte significativo de la responsabilidad. Esto exige otra clase de imperativos. Si la esfera de la producción ha invadido el espacio de la acción esencial, la Moral tendrá entonces que invadir la esfera de la producción, de la que anteriormente se mantuvo alejada, y habrá de hacerlo en la forma de política pública.”.¹⁹³

A su vez, Jonas comenta que la responsabilidad está relacionada con el poder y el saber y en ella ambos mantienen una relación compleja. El poder y el saber eran limitados en otros tiempos, se buscaba ¡el recto hacer en lo actual y concreto y se promovía la virtud ciudadana. “El poder causal es condición de la responsabilidad. El agente ha de responder de su acto: es considerado responsable de las consecuencias del acto y, llegado el caso, hecho responsable en sentido jurídico... El daño causado tiene que ser reparado, y eso, aunque la causa no fuera un delito, aunque la consecuencia no estuviera ni prevista ni querida

¹⁹¹ *Idem.*

¹⁹² Hans Jonas, op. cit., p. 36-37.

¹⁹³ *Ibem.*

intencionadamente. Basta con que yo haya sido la causa activa... esto se da cuando hay una conexión causal con el acto.”¹⁹⁴ En el Derecho Civil es común encontrar en los códigos que el superior, dueño o patrón es el responsable de los actos de sus subordinados (al igual que por ellos recibe ganancias, por ellos debe responder ante alguna afectación) y los padres son responsables de los hijos (por el su inmadurez y tutela nata de los mismos).¹⁹⁵ Claramente encontramos aquí un principio de la responsabilidad civil que tiene un soporte en la responsabilidad moral.

En un segundo momento, este tipo de responsabilidad se mezcló con la idea del castigo (es decir, de lo civil-comercial o mercantil, se utilizó esto para el ámbito penal) utilizando el sentido moral de la responsabilidad sobre el acto causal. Aquí la declaración de culpable cambia y se castiga más el acto que las consecuencias; la sanción se mide por el acto. Para ello, el fiscal investiga el acto mismo (en algunos países con base en la teoría del delito, la teoría del caso y teoría de la prueba), preguntándose cuáles fueron las intenciones, el propósito, motivo o razón del criminal. La pena, vista desde el *ius puniendi*, no sirve para reparar el daño o la injusticia sino para reestablecer el orden moral perturbado. “Es decir, aquí es la cualidad del acto y no su causalidad el punto decisivo del que hay que responder.”¹⁹⁶

El Derecho -Civil y Penal- hacen la distinción de los tipos de responsabilidad. Uno, buscando una compensación obligada por ley (indemnización por daños y perjuicios), el otro, un castigo por la culpa (actualmente ya incluye la reparación del daño). Pero ambos utilizan el concepto responsabilidad y el concepto actos hechos.¹⁹⁷

¹⁹⁴ *Ibidem.*, p. 161.

¹⁹⁵ Hans Jonas da el ejemplo del clavo que falta en la herradura del caballo y la responsabilidad del patrón del obrero para pagar al cliente directo una indemnización ya que se considera responsable de la negligencia de su obrero, aunque a él mismo no le alcance el reproche. “La negligencia del obrero es aquí lo único que en todo caso cabe llamar moralmente culpable, y eso en grado mínimo... La responsabilidad obligada a pagar los daños puede estar libre de toda culpa. El principio de imputabilidad causal queda siempre salvaguardado en la relación por virtud de la cual el superior reúne de manera general en su persona la causalidad de sus subordinados...” Cfr. Hans Jonas, op. cit., p. 161.

¹⁹⁶ *Ibidem.* p. 162.

¹⁹⁷ *Idem.*

Existe, además, un tercer momento de evolución del concepto responsabilidad que se relaciona con la determinación de lo que se ha de hacer: "...yo me siento responsable primariamente no por mi comportamiento y sus consecuencias, sino por la cosa que exige mi acción. La responsabilidad, por ejemplo, por el bienestar de otros no examina sólo los propósitos dados del acto en lo que respecta a su admisibilidad moral, sino que obliga a realizar actos que no tienen otro propósito que ese fin... Aquello por lo que soy responsable está fuera de mí, pero se halla en el campo de acción de mi poder, remitido a él o amenazado por él. Ello contrapone al poder su derecho a la existencia, partiendo de lo que es o puede ser, y, mediante la voluntad moral, lleva al poder a cumplir su deber. La cosa es cosa mía porque el poder es mío y tiene una relación causal precisamente con esa cosa.".¹⁹⁸

El actuar del médico entra en este tipo de responsabilidad. Hay un deber-ser del objeto (la medicina) y un deber-hacer del sujeto (el actuar ético del médico al realizar algún acto). Este tipo de responsabilidad es la que Jonas sugiere estudiar y comparar con los otros sistemas de responsabilidad (responsabilidad orientada al futuro)¹⁹⁹ y tiene como primer mandamiento la existencia de la humanidad. "Existencia de la humanidad significa sencillamente: que vivan hombres. El siguiente mandamiento es que vivan bien. El nudo *factum* óntico de que haya humanidad en general se convierte en mandamiento ontológico para los que no han sido preguntados antes sobre ello: en el mandamiento de que debe seguir habiendo humanidad".²⁰⁰ De lo anterior, se puede concluir que toda tecnología que afecte este mandamiento debe ser prohibida. Y para el caso de la tecnología utilizada para diagnosticar o tratar alguna afectación del cuerpo de la persona, de su salud física y mental, siempre se debe utilizar como una herramienta auxiliar; no como el origen único de la decisión médica sin reflexión de la persona.

Ahora expondremos la visión del filósofo mexicano Carlos Llano Cifuentes (1932-2010). Llano estudió Filosofía en la *Pontificia Università degli Studi San Tommaso d'Aquino*, (*Angelicum*) y en la Universidad Nacional Autónoma de México,

¹⁹⁸ *Ibidem.*, pp. 163-164.

¹⁹⁹ *Ibidem.*, p. 164.

²⁰⁰ *Idem.*

donde conoció a José Gaos, ex alumno del gran pensador José Ortega y Gasset. Carlos Llano fue uno de los fundadores del IPADE, de la Universidad Panamericana y de muchos otros proyectos académicos y empresariales²⁰¹. Se vinculó al ámbito empresarial por cuestiones familiares y ahí explotó al máximo su formación como filósofo. Llano, en muchas ocasiones se jactaba de decir: “Yo le hablo a los empresarios de Filosofía; y a los filósofos, de empresa.”.²⁰² Y efectivamente, las aportaciones más importantes de Carlos Llano Cifuentes al *Management* las trae de su vasto conocimiento de la Ética y la Antropología Filosófica.

Llano pone a la persona como el centro de la empresa. La persona, vinculada a la *paidea*, cuidando su dignidad, su carácter de ente moral, racional, su autonomía para llegar a conseguir sus fines. La empresa -dice Llano- es una comunidad de personas y su convivencia cotidiana *acarrea responsabilidades*. Por esta razón, los directivos deben crear pautas de comportamiento ético y deben usar el Método del caso, es decir, arrancar los conceptos de la realidad; en lugar de aplicarlos rígida y homogéneamente a ella, con ello superaba a la tradicional administración como actividad científica atendiendo a leyes universales y daba la pauta para explicar que la empresa debe tener principios de operación relacionados con la vida del ser humano.²⁰³

El concepto de empresa “personificada”²⁰⁴, es decir, fortalecida con los vínculos que ésta guarda con los seres humanos y poniéndolos en el centro de sus intereses. La empresa es un sujeto con una densa carga de eticidad²⁰⁵. Esa visión derivada de su formación académica y su experiencia profesional -filósofo-

²⁰¹ Para conocer un panorama general de su vida, obra y aportaciones consúltese Héctor Zagal, *Para entender a Carlos Llano*, FCE, México, 2014. Existe también la Cátedra Carlos Llano UP-IPADE, como proyecto de divulgación de la obra de Llano en el campo de la investigación, la difusión y la docencia. En la Biblioteca del IPADE se puede consultar su obra sobre Management y sus textos filosóficos. <https://www.carlosllanocatedra.org/>

²⁰² Cfr. Héctor Zagal, *Para entender a Carlos Llano*, FCE, México, 2014, p. 25.

²⁰³ Presentación del libro: *La enseñanza de la dirección y el método del caso*, en el sitio del IPADE: <https://www.carlosllanocatedra.org/la-ense%C3%B1anza-de-la-direcci%C3%B3n-y-el-m%C3%A9todo-del-caso-rese%C3%B1a>

²⁰⁴ La personificación se da en 4 niveles: del trabajador directivo, del trabajador operativo, del inversionista, y del consumidor o cliente y del vendedor o proveedor. Carlos Llano, *Dilemas éticos...* op. cit., p. 100.

²⁰⁵ Carlos Llano Cifuentes, *Dilemas éticos de la empresa contemporánea*, IPADE-Ediciones ECA, México, 2019, p. 77.

empresario/empresario-filósofo- lo hizo ser un hombre excepcional en sus reflexiones y en la forma de atender las necesidades y problemas de la empresa. Llano conocía perfectamente la labor directiva y la condición humana. Entendía que la inteligencia sumada a la voluntad le daba al ser humano “carácter”²⁰⁶; y que el carácter como instrumento para el cambio de las organizaciones (y, sobre todo, el cambio de la persona) era un elemento fundamental para la acción directiva.

Arturo Picos, lo define así: “Para Llano la acción directiva es una derivación de la complejidad y profundidad del ser humano en la que aquélla tiene su origen. De ahí que la dirección de empresas no se pueda configurar como una mera aplicación de técnicas diseñadas desde un sistema organizativo. Más que por el ejercicio de una función, la dirección de organizaciones está caracterizada por unas cualidades que surgen como el florecimiento del talante personal del director.”.²⁰⁷

El talante directivo concentrado en la voluntad humana y en su uso para lograr acciones directivas debe responder, al mismo tiempo, a principios éticos que marcan las pautas de su proceder y de los resultados. “El proceder ético de la empresa se resume en el desarrollo de las personas que la integran o con las que establece relaciones. El principio personalista de totalidad señala que no debo obtener el beneficio de unas personas mediante el perjuicio de otras.”²⁰⁸.

Para Llano la dirección empresarial no es una actividad científica; es una actividad prudencial que busca “el acierto en cada situación real concreta... la dirección de la empresa está llena de imprevisiones, fuerzas contrarias, resultados sorprendentes, reacciones humanas no racionales y sentimientos”.²⁰⁹ Y, a la vez, “el hombre es, de modo constitutivo, un ser de responsabilidades”²¹⁰ que lo llevan al ámbito de las implicaciones éticas. En el caso de un director en una organización,

²⁰⁶ Cfr. Carlos Llano Cifuentes, *Formación de la Inteligencia, la voluntad y el carácter*, Trillas, México, 2015.

²⁰⁷ Cfr. Arturo Picos, “Prologo”, *Para entender a Carlos Llano*, FCE, México, 2014, p. 10.

²⁰⁸ Cfr. Dilemas éticos... op., cit., p. 101.

²⁰⁹ Cfr. Presentación del libro: *La enseñanza de la dirección y el método del caso*, en el sitio del IPADE: <https://www.carlosllanocatedra.org/la-ense%C3%B1anza-de-la-direcci%C3%B3n-y-el-m%C3%A9todo-del-caso-rese%C3%B1a>

²¹⁰ Cfr. Carlos Llano Cifuentes, *Ser del hombre y hacer de la organización*, IPADE-Ediciones ECA, México, 2018, pp. 334 y ss.

“le acarrea un esfuerzo por infundir conocimientos éticos que sólo pueden clarificarse a partir del estudio filosófico de la personalidad.²¹¹”.

En la época en la que escribió su libro *Dilemas éticos de la empresa contemporánea*²¹², se comienza a considerar como responsables a las personas que no hacen acciones de supervisión dentro de la empresa (inicia la llamada vigilancia de la Ética de la organización), pero ese tipo de responsabilidad no es la que Llano propone. Este tipo de responsabilidad es solo uno de los cuatro campos de responsabilidad que existen dentro de la empresa y se pone como reto explicarlas a lo largo de todo un capítulo de su libro.²¹³

Llano dice: “hay determinados conceptos éticos que han sido marginados de los usos de la administración de empresas (por ejemplo, el concepto de virtud...) ... [pero] no podemos decir lo mismo del concepto responsabilidad... el concepto responsabilidad resulta clave en la actual dirección de las empresas, y es que ésta, en buena parte, consiste en la atribución de responsabilidades y en procurar que los hombres se hagan responsables de las funciones que se les asignan.”.²¹⁴

Esto llevado al campo de la tecnología tiene gran relevancia. Un ejemplo lo encontramos en las reflexiones que hace a la premisa *hight-tech, high touch* de John Naisbitt²¹⁵ y lo que llama integración consciente de la tecnología en nuestras vidas

²¹¹ *Idem.*

²¹² Cfr. Carlos Llano Cifuentes, *Dilemas éticos de la empresa contemporánea*, IPADE-Ediciones ECA, México, 2019, p. 329.

²¹³ Cfr. Carlos Llano Cifuentes, *Dilemas éticos de la empresa contemporánea*, IPADE-Ediciones ECA, México, 2019, p. 329 y ss.

²¹⁴ *Ibidem.*, p. 331.

²¹⁵ Llano leyó a John Naisbitt, autor de los libros: *Megatrends: Ten New Directions Transforming Our Lives*, Warner Books, Inc., 1982 y *High Tech High Touch: Technology and Our Search for Meaning*, Broadway, 1999.

John Naisbitt, junto con Nana Naisbitt y Douglas Philips, hablaban de la relación de los seres humanos con la tecnología (como consumidores de productos, medios y tecnologías genéticas emergentes). Explica el impacto de la tecnología en la vida diaria, en el arte y en el humanitarismo. En *High Tech - High Touch* muestra cómo obtener los mayores beneficios de la tecnología, minimizando sus efectos negativos. Naisbitt trata de responder preguntas como: ¿La tecnología nos libera de las limitaciones del mundo físico o nos hace esclavos de nuestras máquinas? ¿Ahorra nuestro tiempo diario o simplemente crea un vacío que nos sentimos obligados a llenar con nuevas tareas y responsabilidades? ¿La tecnología nos libera de las limitaciones del mundo? En sus últimas apariciones (Conferencia “2020: Visión on print” organizada por Lexmark) se centró en explicar la relación entre las personas y la tecnología y concluir que el Internet no es un fenómeno tecnológico, sino que se trata de un fenómeno sociológico. Naisbitt citó algunos ejemplos de lo que denomina *High Tech – High Touch* y que en resumen se define como la capacidad de algunas tecnologías de apelar a los aspectos más emotivos de las personas.

(identidad, relaciones, comunicación, sentido de comunidad) sumado al fenómeno de la *erosión del ubi* de William Knoke²¹⁶. Llano muestra su pensamiento, crítica y, lo más valioso, ofrece soluciones. Una de ellas es el denominado criterio de proximidad y el “principio de acción subsidiaria”²¹⁷. Por su relevancia nos permitimos hacer una cita completa a las palabras de Carlos Llano:

“Es preciso revalorizar el sentido de la proximidad de las personas, precisamente cuando parece que esta proximidad se ha hecho posible a nivel planetario gracias a los portentosos avances de los medios de comunicación. No estamos seguros de que sea verdadera la premisa *high-tech, high-touch* de Naisbitt, que nos dice que, ***cuanta mayor tecnología, mayor contacto puede lograrse de manera global***. Hemos de percatarnos de que, en no pocos aspectos, la tecnología comunicativa nos aleja de los más próximos... La proximidad exige una *ubicación*. Para ser próximos a alguien, es preciso encontrarnos nosotros en un lugar cercano y que nuestro interlocutor se encuentre también en cercanía nuestra... el ***criterio de proximidad se encuentra comprometido por el avance tecnológico de las comunicaciones, está a la vista en los adictos al internet***: viven en un mundo distante -que se ha hecho próximo- cuando se escapan de lo próximo para navegar en lo distante. Los adictos al internet no se relacionan con hombres de carne y hueso sino con hombres y mujeres *virtuales*. Evidentemente, ***la tecnología no es culpable de este fenómeno***. Ya sabemos que las cosas no son susceptibles de juicios éticos... El criterio de proximidad, que ***me incita a dar mi preferencia de ayuda a quienes están más próximos y a acercarme a quienes han de ser los destinatarios de mi ayuda, es una versión del principio de subsidiariedad, importante eje de la doctrina social cristiana***. Por el principio de acción subsidiaria, ***las entidades superiores no deben de intervenir en el ámbito de la acción de las entidades que les están subordinadas; si intervienen, ha de ser porque esas entidades subalternas no son capaces de cubrir eficazmente el cerco de sus atribuciones, pero han de hacerlo de modo y manera que estas últimas alcancen esa capacidad cuanto antes***. Pues bien, este sabio principio de acción subsidiaria puede generalizarse hacia arriba enunciándolo de la siguiente manera: tanta autoridad como sea necesaria y tanta libertad como sea posible; o bien concretarse hacia abajo con la siguiente enunciación: ***debe atender los hechos la entidad que se encuentre más próxima a ellos y sólo lo hará la más lejana ante la incapacidad -transitoria- de la próxima***... Este criterio resulta importante para la empresa actual: entre la persona y la persona no debe interferir la técnica, si tal interferencia perturba la relación verdaderamente personal, aunque en otros aspectos pueda lograrse una eficacia mayor... ²¹⁸

²¹⁶ William Knoke, *Bold New World: The Essential Road Map to the Twenty-First Century*, Kodansha Amer Inc., 1997. Knoke se ha manifestado sobre la revolución de las comunicaciones y el transporte. Hoy, “cerca es igual a lejos”, por ello, todas las ecuaciones mediante las cuales se había calculado el poder nacional, la riqueza empresarial y la influencia personal deben rediseñarse. En su libro examina la sociedad y sus cambios vertiginosos.

²¹⁷ Según Llano, es un criterio axiológico (conjuntamente con el criterio de extensión y de incidencia), pero no es un criterio de eficacia operativa como sí lo es la colaboración es más eficaz que la competencia, así como la cooperación y persona. Carlos Llano Cifuentes, *Dilemas éticos de la empresa contemporánea*, IPADE-Ediciones ECA, México, 2019, p. 147 y ss.

²¹⁸ Cfr. Carlos Llano Cifuentes, *Dilemas éticos de la empresa contemporánea*, IPADE-Ediciones ECA, México, 2019, pp. 145-147. [El énfasis es nuestro]

De manera vivencial, todos actualmente estamos viviendo con mayor intensidad ese fenómeno de los avances tecnológicos de manera global, la tecnología también nos ha alejado de los más próximos. Pareciera paradójico, pero no lo es. Por un lado, la tecnología nos permite tener contacto con prácticamente cualquier persona en cualquier parte del mundo. Por el otro, nos aleja de la gente que tenemos enfrente, ya que todos, al estar conectados en nuestros dispositivos, estamos a su vez, en otro lugar o espacio. Y la pandemia del COVID -19 incrementó el distanciamiento físico.

El criterio de proximidad y el principio de subsidiariedad son utilizados por Carlos Llano para desarrollar los cuatro campos de responsabilidad. Para fines de nuestra investigación le hemos denominado Teoría de una Moral de la Responsabilidad Empresarial.

La Teoría de una Moral de la Responsabilidad Empresarial de Carlos Llano se convierte en una herramienta de gran ayuda para evaluar la responsabilidad de las empresas por el uso de sistemas inteligentes que generen alguna afectación a Derechos Humanos, ya sea con participación de la voluntad de las personas o sin intervención de ella.

Llano parte del concepto libertad. “Si el hombre es libre -su conducta parte de sí mismo-, es responsable – su conducta ha de atribuirle a sí mismo-.”²¹⁹ Si eres libre de crear ideas y de emprender proyectos, eres responsable de esas ideas y de todo lo que generes. En la iniciativa privada, en el gobierno, en la Teoría Moral y en la Teoría Jurídica de la responsabilidad, esto está claro y nadie lo ha podido objetar. La libertad sin límites ha causado innumerables acontecimientos dañinos para el ser humano y se debe restringir. A esto Llano le denomina responsabilidad del consecuente y tiene que ver, o se traduce, en las categorías de responsabilidad legal que hemos visto. Sin embargo, es un tipo de concepto limitado que no cubre todo lo que hace una empresa (ni los seres humanos). Dice Llano: “La empresa contemporánea entiende la responsabilidad sólo como la condición del hombre que responde -y de ahí el nombre- de las consecuencias de sus actos, concepto que

²¹⁹ Cfr. Carlos Llano Cifuentes, *op. cit.*, p. 330.

resulta en exceso limitativo de esta realidad antropológica tan importante al hombre y la sociedad”.²²⁰

Además de la responsabilidad del consecuente (o responsabilidad de las consecuencias de la acción), Llano ubica la responsabilidad del antecedente (o responsabilidad de los principios), la responsabilidad del congruente (o responsabilidad de los proyectos anteriores de mi vida); y la responsabilidad del trascendente (o de la misión a la que, con mi vida, he sido destinado). “Una moral de la responsabilidad como la que pretendemos desarrollar ahora, ... requiere abarcar estos cuatro amplios campos.”²²¹. Estos son los cuatro elementos de la que denominamos: Teoría de una Moral de la Responsabilidad Empresarial.

Llano hace una advertencia para todos sus lectores: “Restringir el concepto de responsabilidad a la responsabilidad del consecuente... es una restricción grave. Pero, además de esta restricción, la empresa de nuestro tiempo tiene un concepto ambiguo de este mismo tipo de responsabilidad. Esta ambigüedad se capta fácilmente en las consideraciones básicas que pueden hacerse con respecto al liberalismo económico.”²²². Esto se debe tener muy presente hoy, más en el mundo de las nuevas tecnologías ya que de pronto perdemos el origen de la responsabilidad y sus límites; nuestra mente sólo procesa la responsabilidad del consecuente y no los demás tipos de responsabilidad.

Llano muestra dos posturas sobre el tema de la libertad y la responsabilidad. La primera relacionada con el liberalismo económico; la segunda con el modelo de economía socialista. El concepto responsabilidad cambia según el enfoque que se asuma. “El liberalismo económico deja simultáneamente al arbitrio de cada uno la determinación de aquellas consecuencias de sus actos de la que sería responsable. Es importante para una Ética de la empresa, bien enfocada, saber que existe una asimetría entre libertad y responsabilidad. Soy libre de hacer aquello que yo quiero; pero no soy responsable de lo que yo quiera, sino de las consecuencias que derivan de todos mis actos libres.”²²³.

²²⁰ *Ibidem.*, p. 331.

²²¹ *Idem.*

²²² *Idem.*

²²³ *Idem.*

El otro extremo lo encontramos en el concepto socialista de responsabilidad. Aquí el hombre es responsable de todo lo que ocurra en su comunidad. “Si el liberalismo minimiza los alcances de la responsabilidad, la responsabilidad del socialismo es excesivamente maximalista... Nadie puede ser responsable de todo, porque entonces nadie es responsable de nada... El socialista se torna irresponsable por hipertrofia.”.²²⁴

¿cómo encontrar un equilibrio entre estos extremos? Llano crea el concepto de responsabilidad por círculos concéntricos y articula las responsabilidades del ser humano de la siguiente manera:

“...soy responsable, primaria y originariamente, de mi propio ser, y de su fundamento, esto es, de mis virtudes fundamentales...; después soy responsable de mi familia; en tercer lugar, de mi empresa; posteriormente, de las empresas conmigo asociadas; acto seguido, de las empresas de mi ramo; de la economía de mi país, y de la economía del mundo..., pero no en el mismo grado. Aquí se aplica el criterio de proximidad [que ya comentamos previamente]... es decir: soy más responsable de aquello que más próximamente me concierne, y esta responsabilidad mía no ha de ser desplazada por alguien relacionado más lejanamente que yo con los hechos en cuestión... la responsabilidad de o por círculos concéntricos no sólo sirve para la gradación de las responsabilidades (soy más responsable de mí y de mi familia que de mi país), sino que marca también un sentido direccional a mi trabajo: yo no podré influir seriamente en mi país, a menos que me responsabilice de mí mismo, de mi familia y de mi empresa. O, dicho de otro modo, la manera más eficaz para llegar a las responsabilidades más lejanas y difusas es mi fuerza de percusión en los menesteres que me corresponden, más cercanos y concentrados... **La responsabilidad de círculos concéntricos, que no sustituye mis responsabilidades, sino que las gradúa con sentido común y vivido de la realidad de las cosas es un concepto clave para resolver uno de los más graves problemas éticos de las empresas: las consecuencias no intencionales que produce el logro de los objetivos intencionados, cuando nos ocupamos más que de éstos.**”²²⁵

Las empresas, para evitar afectaciones a Derechos Humanos deben ver el tema como parte de una cultura empresarial basada en la Ética de las responsabilidades que todos los seres humanos tenemos y que empieza en mi persona y termina hasta donde me lo permita mi rango de influencia. La responsabilidad bajo este concepto es una responsabilidad compartida por todos los integrantes de la organización y pensando en que forma parte de las medidas preventivas del programa de integridad. Da mejores y mayores resultados que la empresa invierta en un programa de integridad corporativa combinado con un

²²⁴ *Ibidem.*, p. 333.

²²⁵ *Ibidem.*, 334-335.

programa de *Compliance* que sólo se aboquen al cumplimiento de la norma sin un sentido moral organizacional. Además, deben estar conscientes de que su margen de influencia, en muchos casos, es global y por tanto deben ampliar por sí mismas su radio de responsabilidades y no solo quedarse o atender los efectos inmediatos derivados de sus fines específicos.

Llano diría: “hacerse responsables de los efectos consecuentes de estos efectos inmediatos”.²²⁶ “Cuando el empresario tiene los ojos puestos no sólo en el fin primario, sino también en los efectos redundantes del logro de éste, tendría más prudencia para recurrir a la consigna, no pocas veces cobarde, del *sálvese quien pueda*, y exploraría otras posibles soluciones que involucrasen al personal afectado en la búsqueda de sacar adelante la situación, en lugar de desprenderse de él para salvarla.”.²²⁷

En cuanto al segundo tipo de responsabilidad, es decir, la responsabilidad del antecedente esta tiene que ver con que el hombre también es responsable por las razones que toma en sus decisiones. Aunque, el primer tipo de responsabilidad -de las consecuencias- se considere el más amplio y sea en la práctica el utilizado de manera general, existe este tipo de responsabilidad que obliga a los directivos de las empresas y a todo aquel que tome decisiones a dar razones de la decisión que toma; la irresponsabilidad de un director en una empresa es individual y puede convertirse en social.

Llano comenta: “Si un directivo no puede dar las razones de la decisión que toma, es ya irresponsable por ese solo hecho, aunque las consecuencias de la decisión fueran eventualmente favorables... Las razones de mis actos que no necesitan sustentarse en otras, porque dan cuenta de sí mismas, se denominan principios. Por ello, la responsabilidad del antecedente pide que la conducta humana se apoye directa o indirectamente en principios de conducta válidos por sí mismos...”²²⁸

Los directivos de las empresas se enfrentan cotidianamente a la toma de decisiones, si sus decisiones no se pueden justificar incurren en un tipo de

²²⁶ *Ibidem.*, p. 336.

²²⁷ *Ibidem.*, p. 337.

²²⁸ *Ibidem.*, p. 346.

irresponsabilidad. Esto funciona y debe ser un patrón a seguir en la iniciativa privada y en el gobierno, no podemos permitir como ciudadano que la autoridad no justifique sus decisiones ni que las empresas sean omisas en este tipo de responsabilidad. “No se puede admitir la posibilidad de ser un buen director y un mal hombre...”²²⁹

¿Cómo se le puede pedir a los trabajadores que no hagan actos de corrupción si los jefes se llenan los bolsillos de contratos millonarios por actos de corrupción? Es un tema de congruencia que de pronto se olvida. La mejora de la persona a través de la Ética funciona bien en la empresa y algunos principios de la moralidad de Carlos Llano sirven sobremanera para ello. Por ejemplo: “el bien debe seguirse y evitarse el mal”, “no deben emplearse medios moralmente malos aunque los fines sean buenos”, “no deben perseguirse fines buenos que tengan efectos resultantes desproporcionadamente malos”, “ha de considerarse valioso todo aquello que contribuya al desarrollo del hombre”, “hay valores que son objetivos, válidos para toda personas y cultura”, “el hombre debe adquirir las capacidades (virtudes) necesarias para alcanzar una vida lograda, plena o completa”, “obra de tal manera que tu acción pueda erigirse en norma universal de conducta”, “las personas no deben considerarse nunca como medio, sino como fin; entre otras.”²³⁰

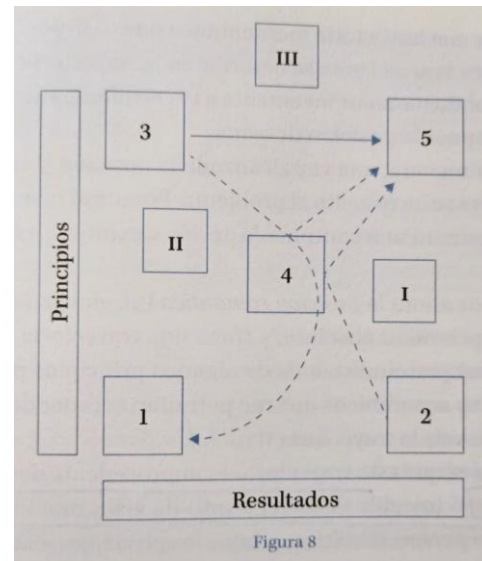
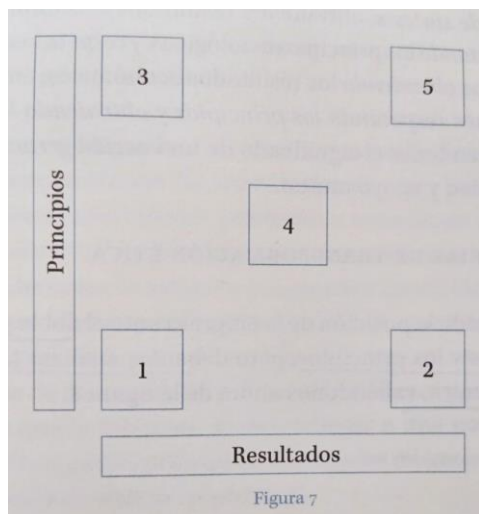
Llano hace una recomendación para medir la responsabilidad del antecedente y utiliza la cuadrícula de Blake para calcular la relación entre los resultados económicos y el cumplimiento de los principios axiológicos. La cuadrícula tiene cinco posiciones que son descritas a continuación:

1. Posición de fracaso: no obtiene resultados económicos ni respeta los principios axiológicos.
2. Posición pragmática: obtiene resultados económicos, aunque no respeta los principios axiológicos.
3. Posición romántica: respeta los principios axiológicos, aunque no obtiene resultados económicos.

²²⁹ *Ibidem.*, p. 349.

²³⁰ *Ibidem.*, diversas páginas.

4. Posición de armisticio: no obtiene todos los resultados económicos posibles porque cumple con algunos de los principios axiológicos, pues si los cumpliera todos, no obtendría ni siquiera estos resultados.
5. Posición de síntesis: obtiene los resultados económicos máximos respetando los principios axiológicos, y respeta los principios axiológicos obteniendo los resultados económicos.²³¹



Fuente: Carlos Llano Cifuentes, *Dilemas éticos...*, op. cit., pp. 351 y 353.

Desde nuestro punto de vista, las empresas deberían utilizar este tipo de responsabilidad y su medición. Muchas de ellas iniciaron en la posición 2 (pragmática), en la que se otorga prioridad absoluta a los resultados, con total menoscabo de los principios. Pero, a la vez, muchas se han dado cuenta que tienen que avanzar hacia la posición 5 (de síntesis)

El tercer tipo de responsabilidad (congruente) tiene que ver con el proyecto de vida de las personas. “La congruencia con el propio modo de ser, con el proyecto de vida, siempre que esté a su vez fundamentada en lo que acabamos de llamar congruencia fundamental (fidelidad a la naturaleza humana); sobre todo, para que

²³¹ *Ibidem.*, p. 351.

los demás sepan a qué atenerse con respecto de nosotros, y podamos inspirar aquella confianza necesaria para ser seguidos por quienes con nosotros trabajan.”.

²³² No se puede contar con un hombre que sigue por norma la ley del gusto, o se limita, como diría Nietzsche, a ejercer su temperamento.

En la responsabilidad congruente el sujeto toma su vida y su destino y se hace responsable de ellos; “procediendo a comprenderse a sí mismo sin imponerse criterios externos de valoración, toma su propia vida en sus manos y ejercitando una fuerte autodeterminación, se hace responsable de sí mismo”.²³³

Finalmente, en cuanto al cuarto tipo de responsabilidad (trascendente), Llano señala que es la que entraña la mayor importancia desde una perspectiva Ética. Cada persona “... no tiene un destino o una misión genérica... sino que cuenta cada una con un texto inédito... cada persona es, pues, responsable de responder a la misión única para la que ha sido destinada”.²³⁴ El ser humano, por tanto, tiene una responsabilidad ante la misión de la empresa y ante la misión personal. Hay responsabilidades insustituibles para el ser humano, como la que tienen los padres con los hijos, los amigos, la familia, “la vida misma... nos crea una situación de insustituibilidad de la que no podemos evadirnos” y esto pasa en la empresa, “en donde nos enfrentamos a eventualidades ante las que no es necesario esperar a que alguien nos asigne un trabajo, en términos de deber, sino que nosotros mismos, anticipadamente, antes de ninguna orden, nos hacemos singularísimamente responsables por las circunstancias, que nos señalan de modo inequívoco nuestro peculiar deber de estado”.²³⁵{

Para efectos de nuestra investigación, pondremos mayor énfasis en la responsabilidad del antecedente y en cómo se toman las decisiones dentro de la empresa. Si son justificadas o no son justificadas. Si se hacen con principios éticos o solo con fines comerciales. En cuanto a la responsabilidad del congruente y el proyecto de vida de las personas, es importante que como organización conozcamos si somos parte de la construcción de ese proyecto de vida o no

²³² *Ibidem.*, p. 357.

²³³ *Ibidem.*, p. 359.

²³⁴ *Ibidem.*, p. 362.

²³⁵ *Idem.*

estamos tomando en cuenta este tema. Es común que en las organizaciones no se tome en cuenta. Nadie es indispensable, se dice por lo general, en lugar de preguntarse cuáles son los planes de vida de nuestros trabajadores y en qué puede apoyarlos la empresa para que los consigan. Y, finalmente, respecto de la responsabilidad del trascendente y sus responsabilidades insustituibles, aunque corresponde a una actividad de grado interior personalísimo, sí se puede participar a través de las áreas de recursos humanos para conocer si hay obstáculo para el cumplimiento de este tipo de responsabilidad. En general todas las responsabilidades están interrelacionadas, y aunque se puede segmentar, el panorama ético funciona mejor concatenando estos temas e incorporarlos a las matrices de riesgos éticos. Saber cómo están tus empleados, socios, accionistas, asociados, sirve para la toma de decisiones, que beneficia todos.

Vale la pena preguntarnos qué tipo de Ética es la de Carlos Llano y cómo serviría para las empresas que están utilizando IA. Para efectos de esta investigación tomamos como base la clasificación de Virginia Dignum. Dignum²³⁶ dice que existen tres tipos de Ética. La Ética basada en el consecuencialismo 1), la Ética basada en la deontología 2) y la Ética basada en la virtud 3).

La Ética basada en el consecuencialismo; sostiene que las propiedades normativas de un acto dependen únicamente de las consecuencias de ese acto. Por lo tanto, un acto se considera moralmente correcto puede determinarse examinando las consecuencias de ese acto: ya sea del acto en sí (utilitarismo del acto, asociado con Jeremy Bentham) o de la existencia de una regla general que requiere actos de ese tipo (utilitarismo de las reglas, a menudo asociado con John Stuart Mill). Esto da lugar a la pregunta de cómo se evalúan las consecuencias. Tanto Bentham como Mill creían que debía prevalecer en las decisiones humanas la regla de la mayor felicidad del mayor número, aunque Mill adoptó una visión más refinada de lo que debería considerarse felicidad. Hay una serie de críticas asociadas con esta noción y han surgido muchas variedades. Hay críticas basadas en la imposibilidad de calcular todas las consecuencias de cada acto para cada persona. También se dice

²³⁶ Cfr. Virginia Dignum, *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*, Springer, 2019, p. 35 y ss.

que el consecuencialismo no pretende ser un procedimiento de decisión, sino un criterio para juzgar las acciones. A pesar de la dificultad de determinar todas las consecuencias reales de un acto (por no hablar de una regla), generalmente hay buenas razones para creer que una acción (o regla) aumentará o disminuirá la utilidad general, y tales razones deberían guiar las elecciones de un consecuencialista. Según el último punto de vista, las acciones se juzgan no en función de las consecuencias reales, sino de las consecuencias esperadas, dada la probabilidad de los diversos futuros posibles. Dada nuestra noción de un agente débilmente ético, el agente mismo recibirá una función de utilidad y elegirá acciones que intenten maximizarla. La elección de la función de utilidad y la forma en que se calculan las consecuencias serán responsabilidad del diseñador.

La Ética deontológica, tiene un elemento clave y es que el valor moral de una acción se juzga por su conformidad con un conjunto de reglas, independientemente de sus consecuencias. Un ejemplo es la Filosofía Ética de Kant. Por ejemplo, en el corazón de la obra de Kant están el segundo imperativo categórico, el concepto de dignidad. Aquí tenemos un principio, el no tratar al ser humano como un objeto sino como un fin en sí mismo. Por tanto, uno debe actuar solo de acuerdo con ese precepto que desearía que se convierta en una ley universal, de modo que las reglas mismas se basen únicamente en la razón. Otra forma de actuación la ofrece la Teoría de la Justicia de Rawls, en la que las normas corresponden a principios aceptables bajo un contrato social adecuadamente descrito. Las reglas para seguir serán elegidas por el diseñador, y el agente moral será un mero seguidor de reglas. Así, el agente moral encarnará sólo una parte muy poco sofisticada de la deontología: cualquier sofisticación reside en el diseñador que desarrolla el conjunto de reglas que seguirá el agente. Los problemas de la Ética deontológica incluyen la posibilidad de conflictos normativos y el hecho de que obedecer una regla puede tener consecuencias claramente indeseables. Son muchas las situaciones en las que se puede considerar incorrecto obedecer una ley del país, y no es difícil imaginar situaciones en las que hay argumentos de que sería incorrecto obedecer también una ley moral. Parte de esto puede manejarse mediante excepciones (que pueden verse como modificaciones que legitiman la violación de la regla general en

ciertas circunstancias prescritas) a las reglas. Sin embargo, prever todas las excepciones es tan imposible como prever todas las consecuencias de una acción. Por esta razón, las leyes a menudo se redactan con cierto grado de vaguedad. Así los jueces pueden interpretar y decidir casos de hechos particulares.

La Ética de la virtud, el enfoque más antiguo de los tres enfoques descritos y se remonta a Platón y Aristóteles. La idea básica aquí es que las acciones moralmente buenas ejemplificarán virtudes y las acciones moralmente malas ejemplificarán vicios. La Ética de la virtud tradicional se basa en la noción de eudaimonia, generalmente traducida como felicidad o florecimiento. La idea aquí es que las virtudes son rasgos que contribuyen a, o son un componente de, eudaimonia. Las alternativas son; Ética de la virtud basada en agentes, que entiende lo correcto en términos de buenas motivaciones y lo incorrecto en términos de tener motivos malos (o insuficientemente buenos), la Ética de la virtud centrada en el objeto, que sostiene que ya tenemos una idea aceptable de qué rasgos son virtudes y lo que implican, y la Ética platónica de la virtud, inspirada en la discusión de las virtudes en los diálogos de Platón. Existe una amplia gama de Ética de la virtud, pero todos ellos tienen en común la idea de que una característica importante de la Ética de la virtud es que reconoce diversos tipos de razones morales para la acción, y tiene algún método (correspondiente a la *phronesis*: prudencia o sabiduría práctica) por considerar estas cosas al decidir cómo actuar.

La Filosofía de Carlos Llano está en este tipo de Ética y es la que tomaremos como nuestro modelo para el estudio de la responsabilidad y el uso de AAMA, haremos una explicación de lo que denominamos Teoría de una Moral de la Responsabilidad Empresarial, la cual va más allá de la Teoría de la Responsabilidad Jurídica y, por tanto, se convierte en una herramienta de gran ayuda para evaluar la responsabilidad de las empresas por el uso de IA basada en Machine Learning – Deep Learning que generen alguna afectación a Derechos Humanos, ya sea con participación de la voluntad de las personas o sin intervención de ella. Analizar el comportamiento ético de las máquinas con base en los valores, principios y reglas que tienen las empresas. Creemos que el camino de la Ética tiene mayores

beneficios, aunque como veremos, es inevitable legislar por parte de los órganos estatales creados para tal efecto.

Si los que procesan los datos, los que desarrollan los sistemas, los que utilizan el sistema, los que lo adquieren y los que en su momento lo auditan, logran entender estas cuatro responsabilidades, y se enseñara esto en las empresas como parte de su política de integridad corporativa o se incluyera en su modelo de gobierno corporativo, se pudiera dar un paso significativo en el fomento al uso de la tecnología y la protección de los Derechos Humanos de las personas. “Tratar de enseñar la dirección de la empresa bajo un concepto cristiano del hombre”.²³⁷ ¿Qué falta además de las recomendaciones de Jonas y Llano? ¿Qué se puede añadir además de esto para concluir el análisis? Creemos que se requiere un ecosistema de IA para la protección de los Derechos Humanos o quizás varios ecosistemas regionales de protección de Derechos Humanos. Este nuevo paradigma se debe construir transdisciplinariamente, ya que de cierta manera está cobrando vida en la Unión Europea con toda la legislación, instituciones y normas de *soft law* que se están creando, pero debe de ir más allá, esos son los ingredientes principales, pero no son los ingredientes suficientes para la construcción del ecosistema.

Si los sistemas inteligentes van a tomar decisiones a semejanza del ser humano, entonces, éstas no deben estar alejadas de la Moral y del Derecho. Las máquinas deben formar parte de un ecosistema de protección social por el uso de IA con un soporte jurídico y con componentes morales tanto para emitir sus decisiones como para la revisión de estas, ya sea por un ser humano o por otro sistema que ayude al ser humanos a proteger sus derechos. Cabe señalar que no todos los sistemas de IA que ayudan a la toma de decisiones son un riesgo para el ser humano, solo aquellos que toman decisiones sin intervención del ser humano o con algoritmos que son considerados cajas negras ya que se tienen ubicadas las entradas (datos) y las salidas (soluciones), pero no lo que ocurre en el trayecto generador de la respuesta o propuesta de solución, lo cual diluye el modelo tradicional de responsabilidad jurídica.

²³⁷ Cfr. Carlos Llano y su relación con San Josemaría Escrivá, fundador del Opus Dei <https://www.carlosllanocatedra.org/blog-carlos-llano/10-aniversario-muerte-carlos-llano-y-50-visita-san-josemaria-escriva-al-ipade>

Evidentemente, existen varias respuestas. Sin embargo, nosotros decidimos tomar el camino de las responsabilidades de la empresa, como comunidad de personas que se unen para ciertos fines y que por esa simple posibilidad de unión de voluntades y de obtención de beneficios se convierten en responsables de su conducta.

Por tanto, ¿qué principios morales deben incorporarse a los sistemas inteligentes? y ¿qué tipo de Moral de la responsabilidad debe incorporarse a las empresas? La respuesta a la primera pregunta está en los Derechos Humanos como parámetro de medición del comportamiento de un sistema inteligente. La segunda pregunta, tiene su respuesta en los postulados de responsabilidad de Carlos Llano Cifuentes y haremos su demostración analizando el caso *IBM Watson Health* en la próxima sección.

Por otro lado, está la pregunta: ¿cómo garantizar a la sociedad que los sistemas inteligentes no tomen decisiones prohibidas por la ley y que a su vez sean respetuosos de los principios morales? Si los seres humanos no pueden determinar por sí mismos que las decisiones o los actos fruto de estas decisiones son conforme a la norma jurídica y no pueden saber si son resultado de la programación humana o de algún humano, entonces se requiere la intervención de otro sistema inteligente construido en un segundo nivel que garantice la no violación de las leyes ni de los principios morales contenidos en cada uno de los Derechos Humanos.

Creemos que se debe pensar en la construcción de un ecosistema de cuidado social para el uso de IA (o quizás varios ecosistemas regionales) y ese ecosistema requiere de un conjunto de elementos que deben permitir ser la puerta de entrada para el desarrollo de políticas, procesos, procedimientos, sistemas y mecanismos de auditoría o vigilancia al interior y exterior de las empresas que generan IA o de los gobiernos que se encuentren regulados y comprometidos a vigilar el respeto a los Derechos Humanos.

Tomamos el concepto de ecosistema de Arthur George Tansley (1871-1955) quien introdujo el término ecosistema y lo definió como un “sistema total que incluye

no solo complejos orgánicos sino también al complejo total de factores que constituyen lo que llamamos medio ambiente”.²³⁸

La problemática es poliédrica y requiere de un estudio profundo bajo el uso de los Derechos Humanos contenidos en la Declaración Universal de los Derechos Humanos (1948) y en los pactos y protocolos facultativos (sin olvidar la futura carta de derechos sobre Inteligencia Artificial que está redactándose en la ONU y en otros sistemas locales de protección de derechos).

Y como propuesta para la construcción de un ecosistema de cuidado social para el uso de la IA consideramos prudente enunciar los ingredientes básicos para su funcionamiento: a) normas, ya sea morales o jurídicas (*de soft law o hard law*); b) agencias reguladoras del gobierno expertas en IA, unas de corte público, responsables de educar, crear políticas, procedimientos, procesos y normas; dirimir controversias, llevar un registro público de algoritmos; certificar y auditar; supervisar el impacto social y sancionar aquella IA que afecte Derechos Humanos; c) Auditores privados, para realizar auditorías de Derechos Humanos a empresas y gobiernos, así como para medir riesgos en esta materia; d) Áreas de *Compliance* ético y legal que se constituyan al interior de las grandes empresas que crean o utilizan IA para la medición de riesgos empresariales (con base en los postulados de Carlos Llano Cifuentes); e) Creación de metodologías para la medición y evaluación del impacto de la IA en los Derechos Humanos; f) Supervisiones periódicas por las agencias reguladoras; g) Tribunales especializados en dirimir controversias que lleguen a sede judicial por temas de responsabilidad algorítmica y afectación a Derechos Humanos; h) Actualización de normas internas; i) Creación de un comité de IA en la estructura de la ONU y sus homólogas regionales, como funciones para casos confirmados de violación de Derechos Humanos por uso de IA, con estructura similar a la de otros comités que existen actualmente²³⁹; j) actualización de planes

²³⁸ Cfr. Tansley AG. 1935. The use and abuse of vegetational concepts and terms. Ecology 16: 284–307. [Reprinted as part of Trudgill S. 2007. Classics in physical geography revisited, Tansley, A.G. 1935: The use and abuse of vegetational concepts and terms. Ecology 16: 284-307.] http://www2.agroparistech.fr/geeft/Downloads/Training/TropEcol/Archives_2011/5.1a.pdf

²³⁹ Actualmente existe el Comité para la eliminación de la discriminación racial; Comité de Derechos Humanos; Comité de Derechos Económicos, Sociales y Culturales; Comité para la Eliminación de todas las formas de discriminación contra la mujer; Comité contra la tortura; Comité sobre los

de estudio en escuelas, universidades y centros de investigación para incluir los nuevos conceptos y usos de la IA y hacer labores de prevención y de divulgación; k) Participación de los organismos empresariales y cámara de comercio en la capacitación y divulgación de estos temas; y, l) utilizar los beneficios de la Ética en la IA. En este último punto vale la pena citar a Virginia Dignum y la descripción que hace de la Ética en estos temas:

A responsibility stance applies ethics to the design of AI in different ways: **Ethics in Design** refers to the regulatory and engineering processes that support the design and evaluation of AI systems as these integrate with or replace traditional social structures. Here the aim is to ensure that for individuals and societies, by anticipating the consequences of the design choices, reflecting upon the problem being solved by engaging all stakeholders, verifying and validating the design and taking appropriate action to ensure social, legal and ethical acceptability of the system. This sponsibility and Transparecy (ART) are at the core of the design of AI systems... **Ethics by Design** is about the ethics of the *behaviour* of AI systems. Work in this field is concerned with (a) the desiderata on the representation, (b) the specification of suitable constraints on system behaviour and (c) the integration of ethical reasoning capabilities as part of the algorithms that determine the behaviour or artificial autonomous systems. **Ethics for Design(ers)** refers to the codes of conduct, the regulatory requirements and the standards and certification processes that ensure the integrity of all actors as they research, design, construct, employ and manage artificially intelligent systems. This is to ensure that they consider the societal impact of their design choices and take the necessary steps to minimize negative impact and dual use of their results. This includes adherence to specific codes of conduct and the definition and use of standards, regulations and certification processes that ensure the integrity of developers, products and services.²⁴⁰

derechos del niño; Comité de protección de los derechos de todos los trabajadores migratorios y de sus familias; Comité contra la desaparición forzada.

²⁴⁰ Cfr. Virginia Dignum, *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*, Springer, 2019, p. 6 y 7. “Una postura de responsabilidad aplica la Ética al diseño de la IA de diferentes maneras: la Ética *en* el diseño se refiere a los procesos regulatorios y de Ingeniería que respaldan el diseño y la evaluación de los sistemas de IA, ya que estos se integran con las estructuras sociales tradicionales o las reemplazan. En este caso, el objetivo es garantizar que, para los individuos y las sociedades, anticipando las consecuencias de las elecciones de diseño, reflexionando sobre el problema que se está resolviendo, involucrando a todas las partes interesadas, verificando y validando el diseño y tomando las acciones apropiadas para asegurar la aceptabilidad social, legal y Ética del sistema. Esta Responsabilidad y Transparencia (ART) son el núcleo del diseño de los sistemas de IA... Ética *por* diseño se trata de la Ética del comportamiento de los sistemas de IA. El trabajo en este campo se ocupa de (a) la desiderata sobre la representación, (b) la especificación de restricciones adecuadas sobre el comportamiento del sistema y (c) la integración de capacidades de razonamiento ético como parte de los algoritmos que determinan el comportamiento o sistemas autónomos artificiales. Ética *para* el diseño (ERS) se refiere a los códigos de conducta, los requisitos reglamentarios y los estándares y procesos de certificación que garantizan la integridad de todos los actores mientras investigan, diseñan, construyen, emplean y administran sistemas artificialmente inteligentes. Esto es para garantizar que consideren el impacto social de sus elecciones de diseño y tomen las medidas necesarias para minimizar el impacto negativo y el doble uso de sus resultados. Esto incluye el cumplimiento de códigos de conducta

El camino es ambicioso y largo, sobre todo porque en México la comunidad jurídica y empresarial esté preocupada de otros temas o por lo menos no logramos ubicar a especialistas que hayan hecho algún análisis dogmático sobre el uso de estas tecnologías y la urgente necesidad de estudiar la responsabilidad de la empresa. Lo cual no significa que nunca nadie haya visto el problema del Derecho y la tecnología²⁴¹; simplemente queremos manifestar que de 2019 a 2021 no encontramos algún experto con formación jurídica o documento que esté investigando y generando documentos para entender estos temas con un enfoque jurídico-empresarial.

También hay que manifestar que es poco común que los estudiosos de la informática intenten vincular sus conocimientos con las aportaciones de la Teoría del Derecho o de la Teoría de la Responsabilidad Jurídica. Ingeniería y Derecho se habían mantenido desvinculados en sus construcciones teóricas y metodológicas, sin hacer una transversalidad de conocimientos; cada uno haciendo aportaciones con su propio método y su propia especialización. Salvo algunas excepciones²⁴²,

específicos y la definición y uso de estándares, regulaciones y procesos de certificación que aseguren la integridad de los desarrolladores, productos y servicios.”

²⁴¹ Quizás el primer autor que intentó poner las bases en el tema de la Ética y uso de tecnología fue Vittorio Frosini, en su libro *Cibernetica, diritto e società*, Edizione di Comunità (Milano, 1968) haciendo un conjunto de reflexiones sobre el impacto tecnológico, los problemas éticos y jurídicos tradicionales. Frosini es considerado el padre del Derecho Artificial en Italia y la creación del concepto “libertad informática” como un nuevo derecho, con el significado de autotutela de la propia identidad informática, es decir, el derecho de controlar (accesar, rectificar o cancelar) los datos personales inscritos en un banco de datos. Existe una edición en español actualizada de Ediciones Olejnik, Santiago, 2019 (ver capítulo V Ética y Cibernética). La cibernética, etimológicamente es la “ciencia de la decisión”. En México, el doctor Julio Tellez Valdés, ha impulsado el derecho informático desde 1983 con su texto *La informática y el Derecho* publicado por el INEGI y su actual obra *Derecho informático*, Quinta Edición, McGrawHill, México, 2021.

²⁴² La Universidad Panamericana está construyendo un proyecto muy interesante en el que participan alumnos y maestros de las facultades de Filosofía, Ingeniería y Derecho. Agradecemos a Karen González Fernández y a Enrique Siqueiros la posibilidad de incorporarnos a esas actividades. Juan G. Corvalán, director del Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires, ha conseguido subir a la Filosofía a la Ingeniería y al derecho con financiamiento internacional para formar especialistas en IA (<https://www.uba.ar/noticia/20195>) La Universidad Autónoma de Chile ha decidido crear un programa para estudiantes en el que unen los conocimientos de las carreras de Derecho e Ingeniería (2019). El programa de “Minor en IA y Derecho” busca desarrollar conocimientos y habilidades “conectadas con la programación, el uso de algoritmos, el diseño de servicios o productos digitales para la resolución de problemas jurídicos y con dominio de sus límites normativos, con relación al plano ético, al de protección a la privacidad y los datos personales y al de seguridad de activos de la información”. Entre las materias que se imparten destacan: Ética e Inteligencia Artificial; Diseño de

las universidades tradicionales mantienen esta dinámica académica. Sin embargo, el desarrollo vertiginoso de la tecnología hace insuficiente estos esfuerzos académicos y se requiere un perfil de profesionistas con formación transdisciplinaria para el análisis y la atención de los problemas que surgen por la creación y utilización de algoritmos de *Machine Learning - Deep Learning*, la responsabilidad algorítmica y la protección de los Derechos Humanos.

La evolución de la IA se ha logrado gracias a la fusión de muchos campos de estudio que inician con la Informática y a Ingeniería Electrónica pasando por las Matemáticas, la Lógica, la Estadística, la Psicología, la Biología, la Lingüística, la Filosofía y el Derecho (con el cual ahora se están construyendo varias normas para la protección de derechos).²⁴³ Si lo entendemos de esta manera, un problema que se genera por el uso del ML o DP puede ser investigado y solucionado con varios enfoques y varios métodos; los métodos a su vez pueden ayudar a encontrar varios problemas y sus posibles soluciones, que no se podrían vislumbrar sin una transdisciplinariedad.

Intentar construir estudios transdisciplinarios de la IA es una realidad que comienza a tomar una dinámica novedosa. Si bien es cierto que nuestro país ya "...tiene una oferta muy amplia de grados, postgrados, sedes, programas, laboratorios y centros de investigación que ofrecen formación en temas relacionadas con la IA"²⁴⁴; que "...la investigación sobre estos temas es prolífica y existen iniciativas colaborativas con el sector empresarial, y también con los

Aplicaciones; Protección de datos personales; Ciberseguridad; Aplicaciones de IA. Cfr. <https://uautonoma.cl/wp-content/uploads/2018/10/Di%CC%81ptico-Minor-IA.pdf>

²⁴³ La informática y la Ingeniería eléctrica determinan cómo se implementa la IA en el software y el hardware. Las matemáticas y las estadísticas determinan modelos viables y miden el desempeño. Debido a que la IA se basa en cómo creemos que funciona el cerebro, la psicología y la lingüística desempeñan un papel esencial en la comprensión de cómo podría funcionar la IA. Y la Filosofía proporciona orientación sobre la inteligencia del ser humano y diversas consideraciones Éticas. Si bien la versión de ciencia ficción de la IA puede ser una posibilidad distante, ya vemos más y más IA involucrada en las decisiones que tomamos todos los días. Desde 1956, la IA ha demostrado ser útil en diferentes campos del saber, impactando la vida de las personas y nuestra sociedad de manera significativa. ML es la herramienta que más impacto ha tenido.

²⁴⁴ Un estudio sobre el estado del arte de la IA en América Latina lo encontramos en: Constanza Gómez Mont; Claudia May Del Pozo; Cristina Martínez Pinto; Ana Victoria Martín del Campo Alcocer (C Minds), *La Inteligencia Artificial al Servicio del Bien Social en América Latina y el Caribe: panorámica regional e instantáneas de doce países*, Banco Interamericano de Desarrollo, México, 2020.

Cfr. https://7be0258a-f55e-4968-87a4-4c067a3d1160.usrfiles.com/ugd/7be025_bbfd0a0224c43a2ad371b87b66c2eae.pdf

gobiernos federal y local”²⁴⁵; y que “se están explorando soluciones innovadoras a retos sociales de manera colaborativa con aquellas organizaciones de la sociedad civil que buscan aprovechar las nuevas herramientas para ampliar su impacto”²⁴⁶; también lo es, que el Derecho, como conjunto de normas y principios que regulan la vida del hombre en sociedad y sus problemas, aún no está presente con la amplitud y profundidad necesarias para generar soluciones. Tampoco ocurre esto con el *soft law* generador de diversos cambios al interior de las empresas.

Un dato curioso es que México es ubicado a nivel internacional como uno de los primeros países que se ha dado a la tarea de crear una “Agenda Nacional de Inteligencia Artificial”²⁴⁷ y existen muchas personas que están involucrándose en estos temas. Sin embargo, la producción de estudios, artículos, ensayos o conferencias aún no se dirige hacia la construcción de un ecosistema de cuidado social para el uso de IA o por lo menos para la creación de un modelo de auditorías algorítmicas con un enfoque basado en los Derechos Humanos. El 17 de agosto de 2021, se presentó la Estrategia Digital Nacional 2021-2024. Los ejes de su acción son: Principio de Austeridad, Principio de Combate a la Corrupción, Principio de Eficiencia en los Procesos Digitales, Principio de Seguridad de la Información, y Principio de Soberanía Tecnológica. Esperemos que el actual gobierno logre implementar diversas acciones y nos muestre sus resultados al término del sexenio.²⁴⁸ Las personas requieren en su vida cotidiana del acceso a la IA y a las demás nuevas tecnologías.

A continuación, haremos el estudio del caso *Watson Health* de IBM, a través de las tres dimensiones expuestas hasta aquí, siendo esto un ejercicio hipotético.

²⁴⁵ *Ibidem.*, p. 96.

²⁴⁶ *Idem.*

²⁴⁷ En abril de 2018 se hizo el documento “Towards an AI Strategy in Mexico: Harnessing the AI Revolution”, patrocinado por la embajada del Reino Unido en México y elaborado por CMINDS (Cfr. <https://www.oxfordinsights.com/mexico>). Posterior se hizo la “Agenda Nacional de Inteligencia Artificial para México” que se ha venido actualizando y fortaleciendo por CMINDS. Puede consultarse en: <https://www.ia2030.mx/> La presentación de la Agenda Nacional de IA2030 de México se organizó por C Minds en el marco de la Coalición IA2030Mx y patrocinado por NVIDIA (2 de septiembre 2020). C Minds es una organización muy activa, a nivel nacional e internacional, con especialistas que están analizando diversos temas y con varios documentos de consulta en: <https://www.cminds.co/reports>

²⁴⁸ <https://www.gob.mx/cedn>

2) El caso Watson Health y la aplicación de las tres dimensiones de la responsabilidad

A) Metodología para la medición de riesgos legales, éticos y tecnológicos.

Antes de entrar a la problematización del caso y a la exposición de nuestros argumentos, consideramos de gran relevancia el uso de un instrumento metodológico que permita sistematizar el análisis de este caso y cualquier otro relacionado con el uso de IA, para lo cual hemos construido una metodología para la medición de los riesgos legales, tecnológicos y éticos basada en un conjunto de indicadores básicos. Estos indicadores fueron seleccionados de los capítulos I; II y III, tomando en cuenta la posibilidad de ser utilizados dentro de una matriz de riesgos, es decir, que sean medibles de alguna manera y auditables. La tendencia mundial, en prácticamente cualquier sector de la industria y los servicios, es la medición de riesgos empresariales, para evitar afectaciones a los recursos humanos, materiales, financieros, tecnológicos o reputacionales.

La metodología busca ser una herramienta que le permita a las empresas incorporar aquellos indicadores de medición básicos tanto en su sistema automatizado como en su matriz de riesgos, y por tanto, detectar y mitigar los riesgos legales, tecnológicos y éticos que producen una afectación o daño a las empresa y sus *stakeholders*.

Tiene tres secciones: La primera, mide los riesgos legales con base en nuestra legislación vigente. La segunda, mide los riesgos tecnológicos que están presentes en las empresas que utilizan IA basada en Machine Learning – Deep Learning, aunque no se encuentren regulados en norma jurídica pero que son ya utilizados en el mundo de la IA basada en Machine Learning – Deep Learning ; y, la tercera sección mide los riesgos éticos que se presentan por la toma de decisiones en la empresa sin un soporte ético.

A lo largo de esta investigación hemos analizado muchos conceptos, principios y hechos ocurridos relacionados con los tres tipos de riesgos que

expondremos a continuación, pero solo elegimos algunos que tienen que ver o dependen de la manera en cómo se dirige y controla a la organización (gobierno corporativo²⁴⁹). Este catálogo se puede ampliar tanto como le sea posible a las empresas y especializarlo según los productos o servicios que ofrecen. No se entra con mayor profundidad a este tema debido a que sobrepasa los tiempos y recursos que tenemos para nuestra investigación, pero valdría la pena desarrollarlos en algún laboratorio de Derecho, Ética e Ingeniería dentro de las universidades.

Para efectos de esta metodología, entendemos por riesgo²⁵⁰ aquel efecto de la incertidumbre (desviación respecto a lo previsto que genera oportunidades o amenazas) sobre los objetivos de una persona u organización (pueden tener diferentes aspectos y categorías, y se pueden aplicar a diferentes niveles); también es una vulnerabilidad a algo, como por ejemplo una externalidad (COVID-19). El riesgo se expresa en términos de fuentes de riesgo, eventos potenciales, sus consecuencias y sus probabilidades. Y, gestionar el riesgo o gestión del riesgo, es aquel conjunto de actividades para dirigir y controlar la organización con relación al riesgo. Existen varios tipos de riesgos (operativos, financieros, de *Compliance*, reputacionales, políticos, contractuales, laborales, del mercado o sector, tecnológicos, de lavado de dinero, de corrupción), para efectos del análisis del caso y derivado de los fines de nuestro trabajo, sólo se hará análisis de los riesgos legales, tecnológicos y éticos.

Nuestra metodología busca establecer parámetros básicos para que las empresas y gobierno que tengan pensado utilizar o ya estén utilizando la IA puedan cumplir mejor con sus obligaciones legales, de tecnología y éticas y, en caso de no contemplar alguna, que se incluya en sus parámetros de medición (deben utilizarlos y añadirlos a la medición de sus productos o servicios, clientes o usuarios, zonas geográficas, canales de distribución, entre otros).

²⁴⁹ Para efectos de esta investigación tomaremos la definición de “gobierno corporativo” contenida en el *Código de mejores prácticas de gobierno corporativo* del Consejo Coordinador Empresarial, el cual lo define como “el sistema bajo el cual las sociedades son dirigidas y controladas”. Consúltense en: <https://cce.org.mx/2021/05/10/codigo-de-mejores-practicas-de-gobierno-corporativo/>

²⁵⁰ Tomada la definición de la Norma ISO 31000: 2018.

La siguiente tabla nos muestra el catálogo de indicadores básicos para la medición de riesgos por el uso de IA. Se puede apreciar cómo las cuatro responsabilidades que contiene la Teoría de una Moral de la Responsabilidad Empresarial permean las tres dimensiones y da soporte al gobierno corporativo de las empresas y, a su vez, se puede apreciar como algunos indicadores de la primera dimensión (privacidad y transparencia) están muy relacionados con la segunda dimensión (a veces pueden no estar regulados y a veces pueden estar parcialmente regulados, pero siempre estarán presentes como un riesgo de IA).

Las dos primeras dimensiones van muy de la mano con el cumplimiento normativo (incluidos los DDHH) denominado *Compliance* (y la responsabilidad del consecuente de Carlos Llano). Por otra parte, la tercera dimensión se encuentra en el ámbito de la integridad corporativa y sumada con las dos primeras constituyen los ingredientes del gobierno corporativo que transversalmente ubica las tres responsabilidades que hemos venido comentando (antecedente, congruente y trascendente) para mitigar los riesgos legales, tecnológicos y éticos. Estas responsabilidades no siempre se toman en cuenta dentro de la toma de decisiones de los directivos de las empresas, pero representan un ingrediente de gran valía para dar soporte a un modelo legal, tecnológico y ético para el uso de la IA.

Lista de preguntas básicas para inculcar una cultura de respeto a las dimensiones legal, tecnológica y ética

Dimensión legal (1era dimensión)	Dimensión tecnológica (2da dimensión)	Dimensión ética (3ra dimensión)
¿Cuenta con un programa de medición de riesgos relacionados con la responsabilidad civil, responsabilidad penal, privacidad y transparencia? ¿Sabe medir los riesgos relacionados con hechos ilícitos y los tipos penales? ¿Sabe medir la responsabilidad por riesgo creado? ¿Cómo previene la reparación o indemnización y el pago de gastos judiciales por algún conflicto legal? ¿Cómo previene la reparación del daño moral? ¿Cómo previene el incumplimiento de alguna cláusula contractual? ¿El uso de IA está regulado por la legislación del país sede de la empresa y qué consecuencias hay por algún incumplimiento?	¿Cuenta con un programa de medición de riesgos relacionados con el tipo de IA? ¿Su programa de medición de riesgos interno mide los riesgos originados por algún posible sesgo algorítmico? ¿Su programa de medición de riesgos mide la opacidad, interpretabilidad, explicabilidad de su sistema de IA? ¿Tiene indicadores para medir la responsabilidad algorítmica? ¿Su organización ha implementado alguna auditoría algorítmica? ¿Cuenta con alguna certificación que avale los impactos positivos/aversos por el uso de IA?	¿Cuenta con algún programa de integridad que esté pensado para medir el cumplimiento ético? ¿Tiene alguna metodología que permita conocer la evaluación de riesgos por impacto ético que produce la IA? ¿Existe algún eticista o filósofo experto en nuevas tecnologías que esté verificando las decisiones que toman los directivos, accionistas y alta gerencia? ¿Ha realizado algún ejercicio de consulta entre sus empleados para saber si la empresa contribuye de alguna manera a que su personal cumpla sus proyectos de vida? ¿Cómo incorpora a sus productos o servicios controles éticos verificables?

Catálogo de indicadores básicos para la medición de riesgos legales, tecnológicos y éticos

Riesgos legales (1era dimensión)	Riesgos tecnológicos (2da dimensión)	Riesgos éticos (3ra dimensión)
1. Responsabilidad civil (DDHH) 2. Responsabilidad penal (DDHH) 3. Privacidad (DDHH) 4. Transparencia (DDHH)	1. Tipo de inteligencia artificial 2. Datos 3. Sesgos algorítmicos 4. Opacidad 5. Interpretabilidad 6. Explicabilidad 7. Responsabilidad algorítmica	1. Responsabilidad del antecedente 2. Responsabilidad del congruente 3. Responsabilidad del trascendente
Compliance (Responsabilidad del consecuente)		Integridad corporativa
Gobierno corporativo basado en la responsabilidad moral (Autor utilizado para este ejemplo: Carlos Llano Cifuentes)		

Fuente: Elaboración propia

Catálogo de indicadores básicos para la medición de riesgos legales, tecnológicos y éticos

Riesgos legales (1era dimensión). Los riesgos legales de una empresa pueden estar en cualquier norma jurídica. Los riesgos legales tienen relación con el objeto de la sociedad, los productos o servicios que genera, los clientes o usuarios, sus zonas geográficas de influencia, los canales de distribución y, sobre todo, los generan los *stakeholders*. Pueden estar en normas civiles, penales, laborales, administrativas, fiscales, tratados comerciales o tratados internacionales de protección a Derechos Humanos. Es muy amplia la gama de riesgos legales. La matriz contempla a las dos principales ramas del Derecho (Civil y Penal) y a los dos principales temas polémicos relacionados con riesgos legales por el uso de IA (privacidad y transparencia), pero como hemos dicho, el usuario de este catálogo de indicadores puede ampliarlo a sus necesidades (por ejemplo, al área médica (como se hizo en este trabajo); al área educativa; de energía; telecomunicaciones, etcétera). Al detectar la norma que obliga a la empresa y a sus empleados, socios o clientes a realizar determinada conducta por el uso de IA (hacer, dar, no hacer) se debe medir el riesgo en su incumplimiento o los factores que pueden influir en el mismo y que están relacionados o tienen su origen en los diferentes tipos de IA. Las medidas que se tomen evitarán caer en el supuesto de la sanción legal. A continuación, se expondrán los indicadores para la medición de riesgos legales:

Responsabilidad civil. Las empresas deben medir los riesgos de responsabilidad civil que surjan por el uso de IA, con base en las reglas de la legislación vigente del lugar en el que tengan su establecimiento y tomarlas en cuenta para evitar una afectación. Las empresas deben crear un programa de medición de riesgos que contemple todos los supuestos de responsabilidad contenidos en el Código Civil que regula sus relaciones de negocios, contractuales o no contractuales y vincularlos con aquellas actividades que están relacionadas con el uso de IA. Para ello, se debe tomar en cuenta el concepto “hecho ilícito” y responsabilidad “por riesgo creado”. Ambas categorías de responsabilidad traen aparejadas la obligación

de la restitución, reparación o indemnización (daños y perjuicios) y el pago de gastos judiciales. Se debe medir también la responsabilidad que surge por daño moral y la responsabilidad que surge de las cláusulas contractuales, que es un tipo de responsabilidad casuística, por lo que en este tema deben revisar sus contratos y las cláusulas que les generan el cumplimiento de alguna obligación. La empresa debe cuidar que el uso de la IA no caiga en los supuestos del hecho ilícito (acción u omisión, detrimento patrimonial, relación de causalidad entre la conducta y el daño o perjuicio, así como la reparación del daño) y, en su caso, que pueda demostrar las excepciones que contiene la ley para la reparación de la afectación (caso fortuito o fuerza mayor, culpa o negligencia inexcusable de la víctima, prescripción, soporte del daño de lo ocurrido por las partes). La empresa debe medir si el uso de la IA puede caer en una responsabilidad por riesgo creado y buscar prevenir la misma con los mitigantes que contiene la propia legislación civil. En muchos casos la responsabilidad civil puede tener relación con acciones u omisiones que regulan las leyes penales (muchos delitos y la responsabilidad penal están relacionados con la responsabilidad civil), por lo que ambos indicadores deben desarrollarse en paralelo. Un programa de responsabilidad civil para la empresa que mide estos riesgos, puede ser un buen complemento para el programa de responsabilidad penal. Y, finalmente, la empresa debe tener presente que existen daños de carácter civil que pueden salirse de control o pueden ser el origen de otra categoría de riesgos por uso de IA que viola Derechos Humanos. Este es un tema nuevo que también se vincula con la responsabilidad penal y el comportamiento de la empresa, el Estado y lo que la teoría denomina justa indemnización por violaciones a Derechos Humanos. Este tipo de riesgos debe medirse tomando en cuenta 1) el tipo de derecho o interés lesionado, 2) la existencia del daño, y 3) la gravedad de la lesión o daño; 4) los gastos devengados derivados del daño moral; 6) los gastos por devengar; 7) naturaleza de la relación jurídica en el marco de la cual tuvo lugar el hecho ilícito; 8) grado de responsabilidad; 9) capacidad económica; y 10) finalidad y objetivo de la indemnización (reparar, prevenir o reprimir).

Responsabilidad penal. Las empresas deben medir sus riesgos de responsabilidad penal por el uso de IA con base en la legislación vigente del lugar en el que tengan su establecimiento. Las empresas deben asesorarse con un especialista que les construya una matriz de riesgos con base en los tipos penales vigentes para saber si alguno está relacionado con el uso de IA. Las empresas deben tener presente que la fuente de la responsabilidad penal radica en las personas y en sus actos u omisiones (volición) y puede ser subjetiva y objetiva. La responsabilidad penal de las personas se determina con el código o la ley penal y la llamada Teoría del Delito, la cual busca sistematizar los elementos positivos y negativos del delito y obtener la culpabilidad de una persona. En muchos sistemas jurídicos, existe la responsabilidad penal de la empresa que se basan en el debido control de su organización (tu eres responsable de tu organización tanto por las ganancias como por las pérdidas y por el uso ilícito de la misma y debes tener medidas de control para garantizar una operación dentro del marco de la ley), como premisa fundamental para imputar o no imputar la responsabilidad de un ente jurídico. Este es un concepto nuevo incorporado en el Derecho mexicano y que funciona también en muchos países de América Latina y Europa. En este tema, las empresas deben verificar que cuenten mitigantes de la responsabilidad penal tales como: a) manuales, de políticas, procedimientos, procesos que regulen la operación de la organización; b) Un código de conducta y un código de Ética que sea difundido en toda la organización y que se demuestre su aplicación real; c) Sistemas automatizados para el control de la organización; d) Auditorías periódicas; e) Sistemas de denuncia; f) Políticas de selección y reclutamiento de personal que busquen evitar la incorporación de personas que pongan en riesgo la integridad de la organización; g) Políticas de manejo de la información (transparencia y privacidad). Las empresas deben distinguir entre aquellas causas generadoras del delito (que surgen de las personas que la integran), de aquellas causas que se les puede atribuir a la propia empresa por hacerse a su nombre, por su cuenta, en su beneficio o a través de los medios que ellas mismas proporcionen. Es muy importante tener presente que cuando muchas personas participan en una actividad es muy común que se diluya la responsabilidad y que no sea rastreable su

imputación, una debida jerarquización de funciones y responsabilidades permite a la empresa evitar alguna responsabilidad por esta omisión. Algunas categorías de responsabilidad penal tienen relación con la responsabilidad civil o con la responsabilidad contenida en leyes administrativas o leyes especiales, como por ejemplo la Ley General de Responsabilidades Administrativas y la Ley General de Salud, por lo que se debe revisar si la empresa tiene algún riesgo que derive de estas normas y, como dijimos en el apartado previo, reconocer que existen reparaciones de daño por violaciones a normas penales que pueden vincularse con reparaciones del daño por Derechos Humanos (lo convierte en un riesgo mayor).

Privacidad. Las empresas deben medir sus riesgos en materia de privacidad por el uso de IA con base en las reglas de la legislación vigente del lugar en el que se encuentren establecidas y tomarlas en cuenta para el manejo de datos personales. El indicador privacidad busca medir los riesgos legales que se generan por el incumplimiento de la legislación de protección de los datos personales al momento de utilizarlos con sistemas de IA. Las empresas mexicanas que tengan negocios en la comunidad europea y las empresas europeas que tengan negocios en México o fuera de Europa, deben de atender las reglas del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos, RGPD), el cual, en su artículo cuarto contiene diversas definiciones relacionadas con datos personales, tratamiento datos biométricos, datos relativos a la salud de las personas, entre otros. La protección de la privacidad está regulada también por instrumentos internacionales como: Declaración Universal de los Derechos Humanos (1948), el Pacto Internacional de Derechos Civiles y Políticos (1966), el Pacto Internacional de Derechos Económicos, Sociales y Culturales (1966) y la Comisión Americana sobre Derechos Humanos (1969). El principal objetivo es desarrollar e implementar un marco que proteja la privacidad y, a su vez, fomente el uso del entorno digital. Este indicador está relacionado con la seguridad de la información, la interpretabilidad y la

explicabilidad de la IA. El principio de privacidad obliga a revisar si el algoritmo protege la información confidencial de los datos de las personas y si es un sistema fiable en sus algoritmos. La Declaración 03/2021 sobre el Reglamento de privacidad electrónica del *European Data Protection Board*, adoptada el 9 de marzo de 2021, habla de hacer cumplir el requisito de consentimiento para las cookies y tecnologías similares, y ofrecer a los proveedores de servicios herramientas técnicas que les permitan obtener fácilmente dicho consentimiento. Esto se convertirá en una regla internacional que seguramente se utilizará por todas las empresas relacionadas con el uso de IA. Este tipo de indicador busca que las empresas mitiguen los riesgos relacionados con un mal uso o un uso equivocado de la privacidad. Los usuarios pueden querer saber qué datos personales está usando un sistema y es posible que deseen controlar ese uso porque sienten que el uso viola su privacidad y esto se convierte en un riesgo para la empresa que se debe de mitigar con controles dentro del funcionamiento del propio sistema de IA. Se deben revisar también las directrices de la OCDE sobre protección de la privacidad y flujos transfronterizos de datos personales tienen por objeto ayudar a los responsables de políticas públicas en cada país para la elaboración de marcos de privacidad. El objetivo de intentar proteger la privacidad puede colisionar con el de evitar sesgos, ya que para entender si una aplicación está sesgada es necesario hacerse de datos personales (género, etnia y algunos otros datos personales de los usuarios). Generalmente, una multa por afectación a la privacidad de las personas es altamente onerosa, por ello se debe revisar las normas y cuidar este tipo de afectación. Este indicador debe ser utilizado conjuntamente con los otros indicadores que miden los riesgos tecnológicos, la opacidad y el orden público tecnológico.

Transparencia. Las empresas deben medir sus riesgos en materia de transparencia de la información por el uso de IA con base en las reglas de la legislación vigente del lugar en el que se encuentren establecidas y tomarlas en cuenta para el manejo de los datos relacionados con el uso de sistemas de IA. El tratamiento de datos personales de toda persona física debe contemplar la recolección, uso, consulta y la medida en que dichos datos son o serán tratados. Existe en este tema el llamado

principio de transparencia que impone diversas obligaciones para las empresas sobre la información de las personas y sus datos personales. Las empresas deben distinguir los posibles destinatarios (o beneficiarios) de la divulgación de los datos para mitigar riesgos que se generan en este tema. Las empresas deben delimitar quién, cuándo, qué y por qué hará la divulgación de datos para evitar riesgos relacionados con la transparencia de la información. Existen aquí riesgos relacionados con el uso de algoritmos de caja negra y cuál es la información que debe revelar la empresa, a quién debe darla y qué tan rápido debe ocurrir la revelación. A esto se le llama transparencia calificada. Las empresas deben utilizar como parámetro o guía los estándares de la *Association for Computing Machinery* (ACM) que en su declaración sobre *Transparencia algorítmica y responsabilidad* ubica varios principios (conciencia, acceso y compensación, responsabilidad, explicabilidad; procedencia de los datos, auditabilidad; y, validación y prueba). Otra norma internacional en este tema, que genera riesgos legales, está en la *Global Privacy Assembly* (GPA) y sus diversas resoluciones que emite (por ejemplo, las transferencias de datos transfronterizas, las cuestiones tecnológicas, la comunicación pública de mensajes de privacidad y las cuestiones de protección de datos dentro de las organizaciones internacionales), éstas se convierten en estándares internacionales que se deben seguir. La transparencia de un sistema de IA está correlacionada con la confiabilidad que genera el mismo en los seres humanos, por ello, cuando no conocemos la trazabilidad algorítmica comienzan las especulaciones sobre los resultados y hasta qué punto fue la mejor decisión o fue la decisión más adecuada al caso concreto. Las empresas deben saber que existen dos tipos de modelos algorítmicos: Uno abierto y claramente interpretable y explicable; y otro, que se considera de caja negra, porque se conocen los datos de entrada y los de salida, pero no cómo se generaron. La empresa debe medir los riesgos de los segundos. La transparencia algorítmica está presente como concepto en el Reglamento General de Protección de Datos de la Unión Europea, el cual requiere que las empresas brinden una explicación *ex post* a las partes involucradas. Se recomienda el estudio de la IA explicable.

Riesgos tecnológicos. Comprenden el análisis y medición de los riesgos derivados por el uso de sistemas de IA basada en Machine Learning – Deep Learning y los conceptos tipo de inteligencia artificial, datos, opacidad, sesgos algorítmicos interpretabilidad y explicabilidad. Esta segunda dimensión, por lo general, no está convertida en norma jurídica, la encontramos en estándares de asociaciones o en recomendaciones de organismos internacionales. Cada uno de estos indicadores debe ser estudiado para verificar si se configura una responsabilidad algorítmica producida por un el concepto de Agente Artificial Moral Autónomo (AAMA), esta es la principal finalidad de esta dimensión, y que, a su vez, permite dividir la responsabilidad que le corresponde a la máquina (responsabilidad algorítmica) de la responsabilidad que les corresponde a los seres humanos. La responsabilidad algorítmica, a pesar de no ser norma vigente en muchos países, es un concepto importantísimo para el Derecho y para medir la responsabilidad de las personas.

Tipos de inteligencia artificial.- Las empresas deben medir los riesgos originados por el tipo de IA que están utilizando. Para efectos de la construcción de este indicador, diremos que existen dos tipos: la IA basada en Sistemas Expertos Decisionales y la IA basada en Machine Learning – Deep Learning. La mezcla de estos tipos de IA es denominada computación cognitiva. Las empresas deben saber que la mayoría de la IA no daña Derechos Humanos ni pone en riesgo a las personas, pero hay un tipo de IA que sí puede generar riesgos relacionados con su uso sin intervención del ser humano y estos riesgos deben ser medidos y mitigados para evitar algún daño contenido en la legislación civil, penal o específica de cada empresa. Se debe conocer los lenguajes de programación y el software que utiliza la empresa. La empresa debe de estar consciente que existen errores de programación que de manera natural están presentes y que se deben de medir para ser mitigados. Se deben analizar las funciones que realiza el sistema o máquina y delimitar el tipo de IA que se va a revisar. En el campo de la IA y la salud es muy frecuente encontrar Sistemas Expertos Decisionales aplicados a cuestiones médicas. Se deben analizar las tres categorías que existen en los sistemas que utilizan ML - DL: 1) “aprendizaje supervisado”, 2) “aprendizaje no supervisado” y 3)

“aprendizaje por refuerzo”. En el caso de DL, se debe de ubicar el tipo de redes neuronales que se está utilizando. Aquí no se debe perder de vista la computación cognitiva y su funcionamiento.

Datos. Las empresas deben medir los riesgos que existen por el uso de datos. El uso de datos debe tener en consideración los siguientes conceptos: a) atributo, b) ente informático, c) campo; d) registro, e) archivo y f) base de datos de la “Lógica de la programación” y también lo relacionado con los “lenguajes de programación” en cuanto a los errores que pueden ocurrir al momento de crear la secuencia de instrucciones y al momento de ser procesadas por la máquina (ya sea con presencia o sin presencia del ser humano). Este indicador busca medir los posibles sesgos algorítmicos que generan los modelos de IA basada en Machine Learning – Deep Learning en sus resultados de predicción y el uso de datos. Aquí se debe de valorar los tipos de datos que se utilizan (estructurados, semiestructurados y no estructurados). El tratamiento de datos personales y la manera de su trato es un claro indicador de riesgo para la empresa (privacidad y transparencia, son riesgos legales). Este indicador tiene una relación directa con las normas jurídicas que regulan la protección de datos dentro de los sistemas jurídicos. La recopilación de datos en línea puede generar riesgos poco predecibles que afecten directamente a los usuarios y, por lo general, los datos siempre tendrán algún tipo de sesgo. Las empresas deben saber que los sistemas de IA pro DL - ML pueden obtener información de las bases de datos o de su entorno y funcionar sin intervención del ser humano para llegar a algún resultado.

Sesgos algorítmicos. Las empresas deben medir los riesgos que tendrán por el uso de la IA y los sesgos algorítmicos. En general, los sistemas de IA basada en Machine Learning – Deep Learning producen sesgos. Desde los datos mismos, su obtención o entrega, en la limpieza de los datos y la selección o creación de variables y en el modelado e implementación pueden surgir sesgos. Aquí se debe evaluar si la empresa, a través de su equipo responsable del uso de la IA, utilizó alguna técnica para mitigar los sesgos algorítmicos. Existen opciones para ajustar los sistemas de IA para reducir los sesgos algorítmicos a través de técnicas para

incrementar el volumen de datos o el ajuste/cambio de indicadores o de los parámetros de muestreo. Los impactos de los sesgos algorítmicos son un riesgo para la empresa, especialmente aquellos que no son explicables. Existe una propuesta para mitigar estos sesgos basada en la gestión de los sesgos relacionados con la intención que tiene la empresa de implementar un modelo de ML; los sesgos que surgen debido a la distribución de muestras utilizadas para la capacitación de ese modelo de ML; y, los que están presentes en muestras de entrada individuales. Aclarando que no existe una solución sencilla ni completa para el problema de sesgos, la combinación de procesos y las auditorías algorítmicas son mitigantes para reducir los sesgos algorítmicos.

Opacidad. Las empresas deben medir los riesgos que tendrán por la opacidad de sus algoritmos. Los llamados “algoritmos de caja negra” han sido cuestionados porque se desconoce su trazabilidad. Este indicador tiene que ver también con el uso de datos y de los sesgos algorítmicos. Las empresas que utilizan algoritmos de caja negra deben estar conscientes que su uso genera riesgos que deben ser medidos para que no se genere un daño legal o reputacional. La empresa debe detectar si la opacidad es el resultado de los propios algoritmos, o es una decisión humana que busca ocultar secretos corporativos, o es una decisión humana derivada de la falta de conocimiento técnico de su equipo de trabajo. La principal bandera roja de la opacidad es el uso de patrones oscuros que genera cierto tipo de algoritmos. La mitigación de este tipo de riesgo se hace a través de los principios interpretabilidad y explicabilidad; y también, de las auditorías algorítmicas para entender su utilización y los resultados de estas. La detección de una opacidad derivada del propio algoritmo tiene que ser atendida con algún otro sistema de IA que pueda dar al ser humano la interpretabilidad y explicabilidad necesarias para no afectar los derechos de las personas. La opacidad algorítmica puede limitar las categorías de responsabilidad y trae consecuencias como la discriminación por motivos de raza, sexo, entorno social, entre otros.

Interpretabilidad. Las empresas deben medir los riesgos que genera la imposibilidad de explicar sus algoritmos a otros seres humanos que pudieran sentirse afectados por un funcionamiento opaco. El uso del principio de la “interpretabilidad” para verificar el funcionamiento de los algoritmos de IA basada en Machine Learning – Deep Learning y su trazabilidad de inicio a fin, en términos que los humanos entiendan, debe convertirse en un deber moral y en una obligación jurídica para la protección de los Derechos Humanos de las personas que todas las empresas deben atender. La empresa debe cuidar que su equipo de trabajo experto en estos temas cumpla los principios para el uso de la IA que se han venido construyendo durante el último lustro. La interpretabilidad como herramienta para explicar la IA, es la articuladora de todos estos principios y tiene como fin construir un sistema de protección de Derechos Humanos para el uso de IA. La interpretabilidad es el principio de principios y, es necesaria para que las empresas protejan a las personas en aquellas actividades relacionadas con comprensión científica, seguridad de la información y, sobre todo, con la toma de decisiones Éticas (justicia, equidad, igualdad, políticas públicas de distribución de recursos; uso de recursos públicos para programas sociales, entre otras).

Explicabilidad. Las empresas deben medir los riesgos por el uso de la IA que genera la imposibilidad de explicar las decisiones de los algoritmos IA basada en Machine Learning – Deep Learning. Si bien, la explicabilidad algorítmica está en algunos métodos de inteligencia artificial (*explanatory artificial intelligence, XAI*), dicha explicabilidad no sirve para explicar a la sociedad o a los usuarios lo que ocurre en el procesamiento de datos con un lenguaje cotidiano. Estamos nuevamente con el concepto opacidad y los “algoritmos de caja negra”. Las empresas se deben preguntar si el sistema de IA o su equipo especializado en estos temas tienen la posibilidad de generar explicaciones “comprensibles para los seres humanos”. Y, en caso contrario, qué mitigantes tiene para evitar una carga de responsabilidad por el uso de IA no explicable. En este sentido, sirve mucho ubicar si los algoritmos tienen explicaciones internas (para expertos de IA) o externas (las exigidas por la sociedad y los usuarios). El Reglamento General de Protección de Datos (RGPD)

de la Unión Europea, crea obligaciones para los procesos de toma de decisiones automáticos y el derecho a una explicación. La empresa medirá sus riesgos tomando como base el tratamiento de la información o los datos que regulan las normas jurídicas, que para el caso, encontramos en el artículo 71 del reglamento en mención, es decir, es el derecho a obtener intervención humana, a expresar su punto de vista, a recibir una explicación de la decisión tomada después de tal evaluación y a impugnar la decisión. Se medirá: a) las circunstancias y contexto específicos en los que se tratan los datos personales, b) si el responsable del tratamiento de los datos utilizó procedimientos matemáticos y estadísticos adecuados para la elaboración de perfiles, c) aplicación de medidas técnicas y organizativas apropiadas para garantizar, y d) si se buscó reducir al máximo el riesgo de error para asegurar los datos personales de forma que se tengan en cuenta los posibles riesgos para los derechos del interesado y se impidan, entre otras cosas, efectos discriminatorios en las personas físicas por motivos de raza u origen étnico, opiniones políticas, religión o creencias, afiliación sindical, condición genética o estado de salud u orientación sexual, o que den lugar a medidas que produzcan tal efecto. Las decisiones automatizadas y la elaboración de perfiles sobre la base de categorías particulares de datos personales únicamente deben permitirse en condiciones específicas. Este indicador busca generar confianza en los usuarios de los sistemas de IA, mostrando sus fortalezas, beneficios y desafíos, y a la vez, sirve para que cualquier parte inconforme del resultado pueda impugnar ante la autoridad el mismo (la vía legal con base en evidencias) y generar una “IA fiable”. Una buena explicación lleva varios ingredientes que se deben incluir en las reflexiones del grupo de expertos que la empresa constituya para medir los riesgos de este indicador.

Responsabilidad algorítmica. - Las empresas deben de conocer si están utilizando algún tipo de inteligencia artificial que genere riesgos de responsabilidad algorítmica y en su caso, mitigarlos con reglas para su uso basadas en los parámetros internacionales de protección a Derechos Humanos. En ocasiones este indicador pudiera estar en norma jurídica vigente, si es así, se debe detectar con precisión en

qué sentido o bajo qué criterio está dicha regulación para poder implementar mitigantes que permitan a la empresa un incumplimiento de los supuestos legales. La norma jurídica pudiera contener obligaciones para la empresa relacionadas con procesos de certificación o requerimientos de auditorías para su uso. En muchos países de América Latina no existe regulación sobre este tema, pero, es recomendable generar buenas prácticas de mitigación de riesgos basadas en los parámetros de regulación de la Unión Europea. Es útil también tomar en cuenta las recomendaciones sobre IA de la UNESCO y de la OCDE ya que ambas tienen un soporte ético bastante útil para la protección de la empresa y de aquellas personas que están utilizando este tipo de sistemas de IA basada en Machine Learning – Deep Learning. Este indicador busca confirmar el requisito de control del sistema o la máquina y evaluar en caso de alguna afectación a alguna persona, si fue responsabilidad del AAMA o de algún ser humano. Este indicador debe de generar resultados que permitan a la empresa, mitigar sus riesgos por el uso de tecnología. Para poder medir la responsabilidad algorítmica previamente se debe de conocer el tipo de IA basada en Machine Learning – Deep Learning que se está usando, los datos, si existe opacidad, si existen sesgos algorítmicos y su relación con la interpretabilidad y explicabilidad. La responsabilidad algorítmica se debe de analizar dentro de la empresa mediante la formación de equipos transdisciplinarios en los que participen abogados, ingenieros y filósofos. La empresa debe saber que es su obligación rendir cuentas y transparentar el uso de sus modelos algorítmicos en determinados casos que tengan que ver con una afectación a los derechos de las personas. La transparencia en los algoritmos es un elemento que permite la rendición de cuentas y que se puede plasmar a través de auditorías algorítmicas.

Riesgos éticos. Comprenden el análisis y medición de riesgos derivados de la operación de la empresa, entendida esta como una comunidad de personas que se unen para ciertos fines (unión de voluntades para obtener beneficios) y por tanto, también es un espacio que genera responsabilidad para sus integrantes. Aquí se utilizarán las aportaciones de la Ética empresarial de Carlos Llano Cifuentes y su Teoría de una Moral de la Responsabilidad Empresarial. En particular aquellos

relativos a la “responsabilidad del antecedente” y la forma en cómo se están tomando las decisiones en la empresa. Si se justifican o no se justifican y si estas justificaciones tienen un soporte en los principios, reglas y valores de la empresa. Las empresas deben de medir los riesgos éticos y mitigarlos con algún tipo de acción que tome en cuenta la responsabilidad del congruente (proyecto de vida de las personas), congruencia con su entorno social, familia, académico; su situación económica (ingresos y gastos); y, la responsabilidad del trascendente (aquellos deberes u obligaciones insustituibles de las personas que deben atender). La empresa debe de pensar como incorpora controles internos permanentes para construir una cultura de la Ética de la IA así como para monitorear el cumplimiento de las responsabilidades con base en la Teoría de una Moral de la Responsabilidad Empresarial de Carlos Llano Cifuentes.

Responsabilidad del antecedente. - Las empresas deben medir los riesgos de IA que generan, directa o indirectamente, sus directivos, gerentes y, en general, de cualquier persona facultada para tomar decisiones. Esta medición debe de hacerse tomando en cuenta los valores, principios y reglas contenidos en sus códigos de Ética y de conducta y en la forma en cómo se justifican sus propias decisiones. Una decisión sin justificación genera un riesgo alto para la organización. Este indicador es complemento relevante para el cumplimiento de las obligaciones jurídicas y, por tanto, para evitar la responsabilidad jurídica de las personas físicas y de las empresas. Se recomienda utilizar como instrumento adicional para la medición de riesgos que derivan de la responsabilidad del antecedente a la llamada cuadrícula de Blake y también se debe tener presente que existen varias aportaciones de la Ética en la IA. Una de esas aportaciones está relacionada con una nueva rama del conocimiento jurídico que explora las preocupaciones técnicas, legales y Éticas que existen en la toma de decisiones con base en algoritmos y su afectación a Derechos Humanos que se conoce como justicia algorítmica. La toma de decisiones algorítmica tiene también que ser analizada por la Ética, lo mismo ocurre con los sesgos algorítmicos y, en conjunto nos llevan a un nuevo modelo de justicia en donde debe de participar la empresa, el ser humano y los AAMA.

Responsabilidad del consecuente y del trascendente. - Este indicador une dos tipos de responsabilidades para que las empresas traten de medir los riesgos éticos provenientes de los proyectos de vida de las personas que se encuentran en toda la estructura de su organización y aquellas responsabilidades que son insustituibles en su vida (familia y amistades). Este indicador se relaciona con las buenas prácticas que existen en materia de gobierno corporativo e indirectamente con temas relacionados con el blanqueo de capitales, sobornos y hechos de corrupción. Lo que este indicador busca es medir la congruencia de que tiene la persona en su forma de vida (ingresos y egresos, entorno familiar y social) y el rumbo o los objetivos que desea alcanzar. La responsabilidad del consecuente y del trascendente requieren una madurez empresarial bastante alta y su medición debe hacerse sin afectar la privacidad de las personas. Es un indicador que se construye de manera voluntaria en el que cada trabajador decide si desea participar o no. La posibilidad de realizar encuestas sobre el clima organizacional y la situación familiar y personal, son herramientas que pueden ser de utilidad para dimensionar los riesgos de este indicador y relacionarlos con el uso de la IA para saber también si se está generando en la organización un ambiente favorable. La empresa, al momento de medir ambas responsabilidades, contribuye a la generación de un ambiente ético que puede traer beneficios para la IA. La aplicación de estas tres responsabilidades permite también determinar riesgos naturales de la tecnología que pueden ser atendidos conjuntamente con el Derecho. A continuación, haremos la problematización del caso *Watson Health* a través del uso del de este catálogo de indicadores.

B) Problematicación de caso *Watson Health* de IBM

Antes de entrar a la aplicación del Catálogo de indicadores básicos para la medición de riesgos por el uso de IA al caso *Watson Health* consideramos prudente explicar cómo el médico atiende actualmente a un paciente de cáncer (caso real, sin IA). Se trata de una persona que fallece después de varios actos médicos realizados con

errores y que se resuelve con la legislación vigente y su doctrina. El caso es el siguiente: Un hombre de 66 años con antecedentes de cáncer de próstata tratado con radiación se presentó a su médico de atención primaria para una cita de seguimiento (de rutina). Se quejó de dolor de cuello y espasmos durante algunas semanas. Se prescribió un antiespasmódico y no se documentaron más antecedentes y exámenes. Tres meses después se presentó para una segunda cita a la clínica. Ahora se quejaba de un dolor que empeoraba en la parte inferior del cuello que se acompañaba de debilidad en el brazo. Las notas del médico documentaron que el paciente tenía dificultades para cargar bolsas desde hace dos meses. Un examen neurológico mostró debilidad de ambas extremidades superiores. Se realizó un diagnóstico de radiculopatía cervical²⁵¹ y se ordenaron radiografías de columna cervical. Se le recomendó al paciente que regresara para seguimiento en tres semanas.

Hay que decir que las quejas de dolor de cuello son comunes en el ámbito ambulatorio y la mayoría están relacionadas con afecciones benignas como distensión muscular y artritis. Para llegar a este tipo de diagnóstico, los médicos a menudo usan el razonamiento heurístico para la toma de decisiones. En este caso, el médico utilizó una heurística de reconocimiento de síntomas y asumió una condición benigna del cuello en la primera visita. Sin embargo, el paciente también tenía antecedentes de cáncer de próstata (una neoplasia maligna que por lo general hace metástasis en la columna), y el médico omite revisar esta información adicional contenida en la historia clínica. Aquí encontramos el primer error del humano: el médico no analizó toda la información para tomar la decisión y esto afectará los niveles posteriores de consulta y evaluación.

²⁵¹ Una radiculopatía es causada por compresión, inflamación y/o lesión de una raíz de nervio espinal en la zona baja de la espalda. Las causas de este tipo de dolor, en orden de prevalencia, incluyen: Hernia de disco con compresión de nervio - es, por lejos, la causa más común de radiculopatía. La estenosis foraminal (el estrechamiento del agujero a través del cual sale el nervio espinal debido a osteofitos o artritis) es más común en los adultos mayores. Ciática, el término que se utiliza comúnmente para describir el dolor radicular que se extiende a lo largo del nervio ciático, describe dónde se siente el dolor, pero no es un diagnóstico real. Por lo general, se llega al diagnóstico clínico a través de una combinación de los antecedentes del paciente (con inclusión de una descripción del dolor) y un examen físico. Los estudios de diagnóstico por imágenes (RMN, TAC, mielograma) se utilizan para confirmar el diagnóstico y, por lo general, muestran el efecto causado en la raíz del nervio. Cfr. <https://www.spine-health.com/espanol/lumbalgia/radiculopatia-lumbar>

El médico debió entender lo que estaba pasando en ese momento y lo que podía ocurrir después (a esto se le llama “conciencia situacional”²⁵²). La predicción de eventos y dinámicas futuras (consecuencias de un proceso metastásico que conduce al síndrome de compresión de la médula espinal en un paciente con cáncer que presenta dolor de espalda y síntomas neurológicos) fueron claramente deficientes y, por tanto, los otros niveles de intervención se perdieron. Esto trajo consigo una ruta de solución incorrecta que implicó solicitar radiografías y concertar una cita ambulatoria futura en lugar de solicitar una resonancia magnética y la admisión inmediata al hospital.²⁵³

En la nueva visita el médico le manda hacer una radiografía de cuello informando al radiólogo que la solicitud se hace porque el paciente presenta dolor. El radiólogo, que desconocía el resto de los hallazgos clínicos del paciente, interpretó la radiografía como aumento de densidad en C6 (vértebra cervical) con posibilidad de metástasis óseas blásticas y recomendó una correlación de gammagrafía ósea nuclear²⁵⁴. Él comunicó personalmente estos hallazgos al médico de atención primaria, quien luego llamó al paciente para que regresara para una exploración.

Cabe resaltar que la información que se le dio al radiólogo fue solo la de dolor, que por sí mismo transmite información ambigua sobre un paciente que tiene hallazgos neurológicos evidentes. Para proporcionar una interpretación óptima de los estudios radiológicos, los radiólogos deben recibir información clínica correcta y pertinente. Aunque la prueba de elección en este caso fue una resonancia

²⁵² Cfr. Endsley, M. R., Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37, 1997, 32–64. <https://doi.org/10.1518/001872095779049543>

²⁵³ La falla en el primer nivel de atención y contacto y en el segundo pudo haber sido por falta de tiempo (esto es algo muy común en los hospitales), debido a la carga de trabajo excesiva, interrupciones y sobrecarga de información. En este caso, el médico estaba trabajando en una oficina que normalmente atiende a 9 a 10 pacientes durante una sesión de medio día, y la mayoría de los pacientes cuentan con múltiples problemas de salud.

²⁵⁴ Una gammagrafía ósea es “una prueba de diagnóstico por imágenes de medicina nuclear que ayuda a diagnosticar y controlar varios tipos de enfermedades óseas. El médico puede solicitarte una gammagrafía ósea si tienes dolor óseo inexplicable, una infección ósea o una lesión ósea que no puede verse en una radiografía estándar. Una gammagrafía ósea también puede ser una herramienta importante para detectar el cáncer que se ha esparcido (metastatizado) al hueso desde la ubicación original del tumor, como la mama o la próstata”. Cfr. <https://www.mayoclinic.org/es-es/tests-procedures/bone-scan/about/pac-20393136>

magnética, el médico siguió la interpretación del radiólogo (que no se basó en los hallazgos clínicos) para obtener una gammagrafía ósea, una prueba quizás apropiada para confirmar la metástasis solo si no había compromiso neurológico en curso. El esfuerzo del radiólogo por comunicar los hallazgos es loable, pero su nivel de comprensión del caso era bastante bajo. Nunca entendió realmente la situación por la falta de información. Los errores en la comunicación entre personas influyen en la toma de decisiones del médico y son la causa de errores médicos que pueden llegar a malas prácticas médicas y a sus consecuencias legales.

Cuando el paciente regresó, se observó que empeoraba la debilidad y la inestabilidad de las piernas. Preocupado por el cáncer de próstata metastásico, el médico de atención primaria se comunicó con un urólogo consultor. El consejo fue realizar un seguimiento en la clínica de urología en tres días. A los tres días, el paciente, en su visita, fue enviado para la gammagrafía ósea. Sin embargo, mientras se realizaba la exploración, desarrolló un dolor de cuello que empeoraba con entumecimiento y parálisis del brazo y la pierna izquierdos mientras estaba acostado en la mesa con el cuello extendido. Fue enviado a la sala de emergencias para una evaluación inmediata. La documentación del médico de urgencias reveló debilidad, ausencia de incontinencia intestinal y de vejiga y reconoció el cáncer de próstata.

Después de la revisión del caso, el médico de urgencias hizo una impresión de radiculopatía de la columna C posiblemente por lesiones blásticas. Discutió el caso con un consultor neuroquirúrgico y se tomó la decisión de realizar una resonancia magnética ambulatoria y un seguimiento en la clínica de neurocirugía en siete días. Al día siguiente, la familia trajo al paciente en una silla de ruedas con empeoramiento de la debilidad y se observó que tenía incontinencia de vejiga. Fue ingresado al hospital y se le realizó una resonancia magnética que reveló que tenía una compresión de la médula espinal metastática en múltiples niveles cervicales. El paciente muere tres meses después.

El síndrome de compresión medular por cáncer es “una de las complicaciones neurológicas con más morbilidad (seguido por metástasis cerebrales). Es considerado como una emergencia oncológica

(independientemente del tipo de tumor, nivel de afectación medular y extensión del tumor primario) porque puede llevar a una pérdida irreversible de la función neurológica e indudablemente a un deterioro en la calidad de vida del paciente sino se realiza un diagnóstico oportuno y se inicia un tratamiento adecuado. La media de supervivencia desde el diagnóstico es de tres a seis meses, por lo cual es de vital importancia detectarlo de inmediato. El tratamiento estándar es la radioterapia; actualmente se han implementado nuevas técnicas de radiación y de cirugía que ofrecen buenos resultados. Se analizan las características clínicas, patológicas, factores pronósticos y manejo de la compresión medular por cáncer, así como las nuevas técnicas diagnósticas y terapéuticas”.²⁵⁵

Además de lo anterior, es importante tener presente que “cuando las células cancerosas se desprenden de un tumor, éstas pueden viajar a otras partes del cuerpo a través del torrente sanguíneo o el sistema linfático... Un cáncer que se ha propagado a un área nueva recibe el nombre de la parte del cuerpo en donde se originó. Por ejemplo, al cáncer de seno que se propaga a los pulmones se le llama cáncer de seno metastásico, y no cáncer de pulmón. El tratamiento también se basa en el lugar donde se originó el cáncer. Si un cáncer de próstata se propaga a los huesos, aún se denomina cáncer de próstata (no cáncer de hueso), y el médico recomendará tratamientos que hayan demostrado ayudar contra el cáncer de próstata metastásico. Asimismo, el cáncer de colon que se propaga al hígado se le llama cáncer de colon metastásico, y no cáncer de hígado.”²⁵⁶

En este contexto, la realización de uno o más actos médicos con errores injustificados, traería como resultado la aplicación de la sanción que la norma jurídica contiene ya sea por negligencia, dolo o impericia.²⁵⁷ Para que haya mala praxis se debe comprobar la culpa del médico en su actuación. De no ser comprobable, pudiéramos caer en un error médico que surge por diversos factores: 1) de tipo institucional, como la burocracia o la carencia de recursos (humanos,

²⁵⁵ María Yicel Bautista-Hernández y Efraín A. Medina-Villaseñor, “Síndrome de compresión medular por cáncer, Gaceta Mexicana de Oncología”, septiembre 2011, pp. 295 a 302. <https://www.elsevier.es/es-revista-gaceta-mexicana-oncologia-305-articulo-sindrome-compresion-medular-por-cancer-X1665920111894309>

²⁵⁶ *Idem.*

²⁵⁷ Ver página 419.

materiales, financieros); 2) una mala calidad de la atención o descuidos en el monitoreo del paciente; 3) la generación de lesiones al cuerpo del paciente que no eran justificadas (con base en el procedimiento que se debió seguir) o la utilización de materiales quirúrgicos inadecuados; o, 4) la contratación de personal poco capacitado para la realización del acto médico con el fin de reducir costo o manipular la facturación. En cualquiera de los escenarios, ya sea culpa del médico o fruto de los diversos factores antes mencionados, las categorías de responsabilidad jurídica tienen opciones para resolver este tipo de problemas. Y en caso de que no fuera así, la interpretación judicial se podría hacer sin mayores complicaciones.

En el sistema jurídico vigente la responsabilidad la asume el hombre (sin ir más a fondo y sin poder responsabilizar a un AAMA). Quizás si se incluyera a *Watson Health*, y se produjera un daño, sería tomado como parte del equipo médico que se utilizó para la atención del paciente (como un bisturí, una radiografía, una jeringa, un examen PCR). No habría responsabilidad para Watson ya que las categorías de responsabilidad están pensadas para los seres humanos. La responsabilidad civil se queda en los médicos (aunque no hayan hecho ningún acto volitivo) y la responsabilidad penal quizás no existiría o existiría por negligencia.

En cualquier supuesto, el sistema Watson sería visto como una herramienta del hombre para actuar juntos y el daño es rastreable o podría dirigirse hacia el equipo médico que se utiliza para esta actividad.

Ahora bien, haciendo un ejercicio hipotético, qué pasa si el sistema jurídico regulara dentro de sus normas el concepto responsabilidad algorítmica, entendiendo por esta aquella que se produce por un Agente Autónomo Moral Artificial (AAMA), que toma decisiones a través de sus propios algoritmos, sin la participación de la voluntad o decisión de un ser humano y que también existiera la justicia algorítmica²⁵⁸. Ubicar a la responsabilidad siguiendo estos nuevos conceptos. Entremos ahora a la problematización del caso y sus argumentos se harán con estos y otros conceptos:

²⁵⁸ Ver página 445.

En 2017 el Centro de Tecnología Cognitiva del Hospital Ángeles, anunció que estaba utilizando *Watson of Oncology* para atender a sus pacientes con cáncer²⁵⁹. El Hospital Ángeles creó el Centro de Tecnología Cognitiva y, a través de este Centro se atiende a los pacientes de cáncer mediante el sistema *Watson Health*. El Hospital Ángeles anunció que incorpora dicha tecnología ya que fue creada por especialistas de IBM y del *Memorial Sloan Kettering Cancer Center*. El Hospital Ángeles lo anuncia como un sistema cognitivo que permite a los médicos oncólogos, brindar tratamientos personalizados a los pacientes con cáncer. Entre las características que da a conocer a los pacientes destacan las siguientes: el sistema selecciona las mejores opciones de tratamiento en oncología, busca y obtiene en minutos respuestas precisas. Esto lo logra a través de comparar millones de datos contenidos en cientos de libros de medicina en miles de publicaciones de revistas científicas de casos clínicos e incluso de notas médicas de reconocidos especialistas en todo el mundo. *Watson Health* compara el caso particular con otros casos clínicos de pacientes desde el *Memorial Sloan Kettering Cancer Center*, por lo que las opciones sugeridas serán siempre basadas en evidencia.

El área de *Compliance* del hospital no se pronuncia ni sobre su contratación, ni sobre su uso y los posibles riesgos, lo ve como un software más que se está adquiriendo y como es una propuesta aprobada por unanimidad por la Asamblea General de accionistas decide omitir su intervención y permite que se inicie operaciones atendiendo únicamente al visto bueno del área legal, financiera y comercial. Todas coinciden en que al tener el respaldo del *Memorial Sloan Kettering Cancer Center* se mitiga cualquier riesgo legal para la empresa, por lo que replican los argumentos vertidos en sus catálogo de operación y aplicando la Ley General de salud y la NOM de dispositivos médicos (Watson se considera como un dispositivo médico y consiguen la certificación de la NOM sobre dispositivos médicos).

A los pocos meses de iniciar operaciones, un paciente es informado de la existencia de esta tecnología y decide hacer uso de ella. Para ello, llena un contrato

²⁵⁹ Puede consultar el siguiente video de divulgación de *Watson of Oncology*: <https://www.youtube.com/watch?v=SUR8K-3x7M0&t=94s>

de prestación de servicios médicos, en el que autoriza que su tratamiento sea bajo la guía de *Watson Health* y el médico especialista en oncología que se le asigne para su revisión y valoración. Para determinar el diagnóstico del tipo de cáncer y el tratamiento a seguir se le hacen varios estudios que son incorporados al sistema *Watson Health*. El médico selecciona el caso del paciente que quiere consultar dentro del sistema *Watson Health* y comparte los datos del paciente con el Centro de Tecnología Cognitiva del Hospital Ángeles donde son procesados a través de *Watson Health*. El oncólogo tratante recibe a los pocos minutos un correo de respuesta con toda la información encontrada por *Watson Health*, con las opciones de tratamiento más relevantes categorizadas en función de su efectividad para este paciente. *Watson Health* pide más información y sugiere realizar al paciente estudios adicionales, para aumentar el porcentaje de certeza en las opciones obtenidas y, adicionalmente, añadir esos datos y vincularlos con nuevas búsquedas de información. El médico atiende la recomendación de *Watson Health* y envía los datos requeridos. Días después, *Watson Health*, hace llegar la respuesta al Centro de Tecnología Cognitiva del Hospital Ángeles y éste a su vez la manda al médico, con el diagnóstico y el plan de tratamiento personalizado para el paciente. El oncólogo, sin tomar ninguna decisión o valoración adicional, aplica el tratamiento de *Watson Health* al paciente y el cáncer es tratado con radiación. El paciente es dado de alta y se retira del hospital.

Un año más tarde, el paciente se presenta a una consulta de rutina. Se queja de dolor de cuello y espasmos durante algunas semanas. El médico de primer contacto le receta un antiespasmódico y no documenta más datos. Sólo hace el reporte al médico consultor de segundo contacto para que ingrese el caso a la base de datos de *Watson Health* ya que por protocolo interno del hospital todos los casos de pacientes de oncología deben ser capturados en el sistema *Watson Health*.

Tres meses después se presenta para una segunda visita de revisión. Ahora se queja de un dolor en la parte inferior del cuello que se acompañaba de debilidad en el brazo. El médico de primer contacto documenta que el paciente tenía dificultades para cargar bolsas desde hace varios meses antes. El médico diagnostica radiculopatía cervical y se ordenaron radiografías de columna cervical.

Se le recomendó al paciente que regresara para seguimiento en 3 semanas y de nueva cuenta se reporta al médico consultor de segundo contacto para que se incorpore a la base de datos del sistema *Watson Health* y se emita una opinión.

En la tercera visita, el médico le manda hacer una radiografía de cuello informando al radiólogo que la solicitud se hace porque el paciente presenta dolor. El radiólogo, que desconocía el resto de los hallazgos clínicos del paciente, interpretó la radiografía como aumento de densidad en C6 (vértebra cervical) con posibilidad de metástasis óseas blásticas y recomendó una correlación de gammagrafía ósea nuclear. El radiólogo le comunicó personalmente estos hallazgos al médico de atención primaria, quien luego llamó al paciente para que regresara para una exploración.

En la cuarta visita, el médico tratante observa que empeoraba la debilidad y la inestabilidad de las piernas. Preocupado por el cáncer de próstata, el médico se comunica con el médico consultor de segundo nivel responsable de Watson pidiendo indicaciones sobre el caso. El consejo del médico de segundo nivel fue programar un seguimiento en la clínica de urología dentro de tres días y pedirle a Watson que hiciera nuevamente una valoración del caso y de régimen de tratamiento específico, identificando datos sobre casos similares. Se le pide también buscar evidencias y compararlas para detectar las características relevantes para el diagnóstico y tratamiento. Se le pide revisar los resultados de las pruebas de laboratorio existentes en su expediente clínico y la búsqueda de notas médicas y artículos en revistas especializadas para conocer lo más actualizado del tema.

Tres días después, el paciente se presenta en la clínica de urología para la gammagrafía ósea. Sin embargo, mientras se realizaba la exploración, desarrolló un dolor de cuello que empeoraba con entumecimiento y parálisis del brazo y la pierna izquierdos mientras estaba acostado en la mesa con el cuello extendido. Fue enviado a la sala de emergencias para una evaluación inmediata. El médico de urgencias reveló debilidad, ausencia de incontinencia intestinal y de vejiga y reconoció el cáncer de próstata y lo reportó al médico tratante y al médico consultor de Watson.

Después de estabilizar al paciente y de hacer la revisión del caso, el médico de urgencias hizo una impresión de radiculopatía de la columna C posiblemente por lesiones blásticas. Discutió el caso con un consultor de Watson y con el médico tratante. Watson emite su informe y llega a la conclusión de que se debe hospitalizar al paciente e iniciar un tratamiento para una enfermedad rara de columna. El médico tratante y el médico consultor de segundo nivel que controla a Watson inician el tratamiento. El paciente fallece días después.

La familia del paciente, molesta por el fallecimiento y tomando en cuenta que se les prometió el mejor tratamiento existente a través de *Watson Health*, inicia una demanda por daños y perjuicios y una denuncia penal en contra de los médicos y del hospital.

El abogado de los médicos, que a su vez defiende al hospital, hace un planteamiento argumentativo que lleva hacia la revisión de las categorías de responsabilidad jurídica vigentes y afirma que ni el hospital ni los médicos son responsables por lo ocurrido ya que Watson le dio al paciente el mejor tratamiento y que lamentablemente hubo un deterioro de su salud que era inevitable y natural.

El médico tratante, tanto en el juicio civil como en la carpeta de investigación, afirma que reportó el caso al médico supervisor y administrador de Watson porque así dice el protocolo de actuación médica del hospital. Él afirma que Watson es un sistema que fue comprado por el hospital para auxiliar a la toma de decisiones médicas y que se les pidió a los médicos reportar todas las consultas; lo cual lo hizo desde la primera cita del paciente, que si bien no tomó en cuenta el antecedente del cáncer de próstata, sí informó al médico consultor de segundo nivel para que incorporara el caso al sistema Watson y que el sistema, a decir de la capacitación que se les dio a todos los médicos, debió de emitir una alerta preventiva del caso ya que los datos de la persona estaban ya capturados por el tratamiento que se le dio para curar el cáncer de próstata. El sistema no alertó a los médicos, tampoco dio un diagnóstico acertado ya que el paciente murió por cáncer de próstata metastásico (y no por una enfermedad rara de columna).

El médico consultor y administrador del sistema Watson, argumenta que fueron capturados los datos al momento de ser reportados por el médico tratante y

que el sistema no emitió ninguna alerta para avisar del perfil clínico del paciente. Reconoció también que el diagnóstico fue erróneo y que no se atrevió a oponerse a su recomendación, siguiéndola al pie de la letra con el médico tratante. Ambos médicos reconocen que tuvieron dudas, pero decidieron seguir las recomendaciones del sistema ya que se les había explicado que Watson tenía la habilidad de hacer búsquedas de casos médicos en las bases de datos de medicina más importantes y actuales que existen en el mundo bajo la supervisión del *Memorial Sloan Kettering Cancer Center*. Además, los médicos no reconocen su responsabilidad bajo el argumento que ellos no tomaron ninguna decisión. Fue el sistema el que procesó toda la información, generó el diagnóstico y el tratamiento erróneo.

El hospital contesta que tampoco le corresponde pagar el daño a la familia porque sus médicos fueron capacitados y fueron entrenados para resolver con base en el protocolo de atención que recomendó IBM para el uso del sistema Watson y que siguieron todas las recomendaciones preventivas, de capacitación y de supervisión de la información; que, si bien los médicos tuvieron errores, su gravedad no encuadra en una negligencia médica ni en algún hecho ilícito o delito.

El abogado del paciente que falleció exige al hospital que demuestre que tuvo el debido control de su organización y que se hizo todo lo que estuvo en sus manos para evitar esta falla o cualquier mal funcionamiento. El hospital únicamente entrega el contrato de prestación de servicios que firmó con IBM *Watson Health*, todos sus anexos técnicos y el programa de capacitación que recibieron los médicos y demás personal responsable de la operación de *Watson Health*. El área legal conjuntamente con el área de *Compliance* comienzan a revisar el modelo de gobierno corporativo y las medidas de debido control en la organización para cuidar su uso y el cuidado de los pacientes y se da cuenta que fueron omisos en su aplicación.

El abogado del hospital manifiesta que la responsabilidad es del proveedor del servicio, es decir de IBM, y no del hospital y le pide al fiscal que un perito emita un dictamen sobre el funcionamiento de *Watson Health* y la imposibilidad de hacer responsable a los médicos que participaron ya que no se comprueba el acto volitivo.

Sin embargo, de los peritos que están en la lista del Tribunal Superior de Justicia de la Ciudad de México, al ser consultados sobre la posibilidad de realizar alguna auditoría algorítmica al sistema *Watson Health*, manifiestan que no cuentan con los conocimientos para la elaboración de un dictamen de responsabilidad algorítmica. Ni ante el Consejo de la Judicatura Federal ni ante el Tribunal Superior de Justicia de la Ciudad de México. De la revisión al listado de peritos se encontraron unos en responsabilidad médica, en medicina, en urgencias quirúrgicas, en medicina forense, en ciencias de la informática, pero no especializados en IA basada en Machine Learning – Deep Learning enfocados al área de la salud (oncología).

El abogado del hospital pide al fiscal, ante la falta del perito experto, que se autorice la contratación de una empresa especializada en auditorías algorítmicas (*compliance* tecnológico) para la revisión del sistema autónomo de IA. Se contrata a la empresa para llevar a cabo una auditoría mediante la problematización del caso basada en dos preguntas: 1) Que diga si existió algún sesgo algorítmico que haya producido una falla o el mal funcionamiento de *Watson Health*; y 2. Que diga si se puede determinar la responsabilidad algorítmica de *Watson Health*.

Para dar respuesta a ambos cuestionamientos, la empresa crea una metodología basada en un Catálogo de indicadores de medición de riesgos legales, tecnológicos y éticos (*Compliance and Ethics Management System*) vinculado a un sistema que interconecta estos módulos para intercambiar y producir información que permita determinar los impactos adversos a derechos humanos y otro tipo de daños. Dicho documento fue estructurado con tres módulos. El primero, tiene que ver con la revisión de toda la legislación que regula a la empresa (interna y externa) para ubicar su funcionamiento como organización responsable de las actividades que se realizan diariamente y el uso de la IA. Se busca detectar riesgos legales relacionados con la responsabilidad civil, penal y lo relativo a la privacidad y transparencia de la información.

El segundo módulo va enfocado a la revisión de aquellos conceptos relacionados con los riesgos tecnológicos para detectar si el sistema, en su operación, es responsable de la mala decisión y si existieron sesgos algorítmicos y opacidad que impida reconocer la responsabilidad algorítmica. Y finalmente, el

tercer módulo analiza las responsabilidades éticas de la empresa y evalúa su programa de integridad corporativa. Al aplicar la metodología se llega a los resultados siguientes:

Fase 1: Riesgos Legales: La empresa de auditoría algorítmica solicita el acta constitutiva y una reunión con todos los responsables de las áreas que la integran para conocer su estructura organización, funciones, modelos de negocio, programa de *compliance* y de integridad corporativa, metodología para la evaluación de riesgos. Realiza a la par, un estudio de la legislación vigente para detectar las normas que debe cumplir la empresa y ubicar la responsabilidad por el uso de la tecnología, desea conocer si la legislación mexicana clasifica a *Watson Health* como un software, un dispositivo médico o un sistema autónomo con regulación específica. La empresa también pide conocer la legislación que regula los temas de privacidad y transparencia y revisar su archivo documental para ver el tratamiento de datos en la organización, en particular, aquellos relacionados con los pacientes que utilizan *Watson Health* e inicia la revisión para detectar si es posible fijar la responsabilidad en alguno de los médicos.

Resultados de la revisión de la legislación civil (responsabilidad civil): De la revisión hecha a la legislación civil (Código Civil Federal) se encontraron dos categorías de la responsabilidad. La primera funciona bajo la premisa del hecho ilícito y la segunda bajo la premisa de la responsabilidad por riesgo creado, ambas categorías de responsabilidad tienen como origen y destino al ser humano y buscan una reparación o indemnización y el pago de gastos. También existen las categorías de responsabilidad contractual y extracontractual. Por la naturaleza del acto jurídico se llega a la conclusión de que encuadra en la responsabilidad contractual. El contrato que firmó el paciente fallecido hace referencia a las reglas de responsabilidad del Código Civil Federal, su clausulado no se encuentra relacionado con alguna falla en el diagnóstico del tratamiento de *Watson Health* que permita una reparación del daño causado por la propia máquina/sistema. Únicamente habla de negligencia médica, dolo y culpa del ser humano.

También se hace un intento por encuadrar la responsabilidad en el riesgo creado por uso de un artefacto peligroso. Sin embargo, tampoco es factible exigir la responsabilidad por riesgo creado ya que no hay participación de algún ser humano en la toma de decisiones que afectaron al paciente, fue el propio sistema el que emitió el diagnóstico equivocado y el tratamiento a seguir.

Respecto a la prevención de riesgos y su mitigación, derivados del uso de la IA y la responsabilidad civil, se preguntó al hospital sobre sus controles internos y sobre la medición de riesgos que deriven del uso de *Watson Health*. El hospital informa que tiene un gobierno corporativo y un *Chief Compliance Officer* enfocado al tema de los servicios de salud, farmacéutica y riesgos penales, pero no atiende los riesgos derivados del uso de la IA que surgen de un contrato de prestación de servicios que se firmó con IBM, ni con ningún otro tipo de empresa con la que tiene relaciones comerciales.

El hospital manifiesta que no existe un programa de responsabilidad civil ni tampoco ha pensado en la revisión de sus normas internas ya que desconocían de este tema. Concluyen que si hubo una afectación o daño al paciente y que la gravedad del diagnóstico que generó *Watson Health* puede devengar gastos al hospital pero que no existe en la legislación mexicana un supuesto que regule la responsabilidad algorítmica por el uso de agentes autónomos que toman decisiones por seres humanos.

Resultados de la revisión de la legislación penal (responsabilidad penal).- De la revisión hecha a la legislación penal se encontraron dos categorías de la responsabilidad. La primera funciona bajo la premisa de la responsabilidad subjetiva y la segunda bajo la responsabilidad objetiva. También se encuentra regulada la responsabilidad de las personas jurídicas. Esta última, toma relevancia para la empresa auditora ya que, al momento de la revisión del marco normativo relacionado con la legislación civil, se manifestó que existía un *Chief Compliance Officer* y que era el responsable de llevar las acciones para la mitigación de riesgos derivados de los servicios de salud, farmacéutica y riesgos penales. Ni el titular del área legal, ni el titular del área de *compliance* están capacitados en estos temas; tampoco conocen este tipo de tecnología y los riesgos que puede producir. El último

reconoce que no aplicó su manual para prevenir los riesgos por la firma de contratos que están sujetos a una revisión específica; no tiene compliance tecnológico.

En la reunión de revisión de las acciones que ha realizado el área legal y de cumplimiento para la mitigación de riesgos por el uso del sistema *Watson Health*, no se encontró ningún indicador ni ningún manual o política que mencionara las acciones a seguir ante una afectación a un paciente por el uso de un sistema automatizado para la toma de decisiones médicas como lo es *Watson Health*; tampoco se encontró alguna persona o política interna para atender este tipo de riesgos. Se tienen controles de cumplimiento normativo basados en la conducta del ser humano, y un equipo de *compliance* que verifica el debido control de la organización en temas de legislación penal, civil, laboral y fiscal, pero no para mitigar riesgos que surgen por contratos para el uso de nuevas tecnologías ni se miden los riesgos al momento de utilizar esas tecnologías para la mejora de la salud del paciente.

En la reunión también se expone el programa de cumplimiento y la matriz de riesgos que mide algunos tipos penales como el de robo, fraude, abuso de confianza, operaciones con recursos de procedencia ilícita, hechos de corrupción, abandono de personas, entre otros, pero no existe algún indicador que mida o se correlacione con el uso de IA basada en Machine Learning – Deep Learning y el sistema *Watson Health*. Al momento de exponer las acciones que se realizan para la mitigación de riesgos se aprecia que el hospital no cuenta con manuales de políticas y procedimientos que regulen el uso de la IA y la toma de decisiones por un agente autónomo moral artificial. Tampoco cuenta con un código de Ética y un código de conducta que incorporen el uso de esta tecnología. Si bien existen auditorías anuales, se han venido solventando las observaciones de manera lenta e incompleta y no se toca el tema del uso de la IA como un riesgo que se mida y tenga alguna acción de mitigación. Esto se debió de incorporar a su metodología para la evaluación de riesgos desde el momento en que se hizo la contratación del sistema automatizado con IBM.

El auditor recomienda a la empresa actualizar todos sus instrumentos internos de *Compliance*, así como su matriz de riesgos para incluir aquellos que

derivan del uso de la tecnología y, en particular, de sistemas autónomos de apoyo médico y se le explica al *Chief Compliance Officer* que al tener el hospital relaciones comerciales con empresas europeas, debe de cumplir con los cambios legales que están vigentes en la Unión Europea y que se relacionan con el uso de IA que puede afectar Derechos Humanos, siendo el sector salud uno de los que tiene medidas de control con mayor rigor, además de exigir la medición de riesgos, se debe de certificar y auditar al sistema de IA que se pretende utilizar para la atención de pacientes. En caso de no realizar esto, pudiera verse sujeto a un procedimiento de sanción y el cierre de operaciones con la Unión Europea.

Resultados de la revisión de la legislación relacionada con la privacidad y la transparencia. – Aunque el tema de la responsabilidad penal y civil por un Agente Artificial Moral Autónomo queda descartado al no existir la regulación de la responsabilidad algorítmica, consideramos importante mencionar que los temas de privacidad y transparencia pueden generar responsabilidad a la empresa que está utilizando el sistema de IA y se pueden convertir en un problema serio que atender ya que existen normas que ya la contemplan y que provienen de la legislación europea.

El hospital cuenta sí cuenta con una política de privacidad y de transparencia relacionada con el uso de la IA. Atiende los criterios y buenas prácticas de la Unión Europea y su reglamento sobre la privacidad y el uso de IA. Muestra las evidencias legales que tienen que ver con el uso de la información capacitada y especializada en temas de tecnología y si existe dentro de la organización alguna persona o especialista que atienda el principio de transparencia de la información con base en la legislación europea. El hospital informa que el responsable de cumplimiento es quien supervisa estos temas y que realizará una exposición de las medidas que ha tomado el hospital para mitigar los riesgos por el mal manejo de la información de pacientes, pero que no tiene ninguna regla o procedimiento para regular la IA y mitigar los riesgos algorítmicos. El *Chief Compliance Officer* desconoce que existía esa obligación ya que la tecnología utilizada es considerada como un software más, y como tal, no requiere alguna acción de cumplimiento; además IBM lo vendió con

diversas certificaciones internacionales. También manifiesta desconocer si *Watson Health* u otro modelo algorítmico es interpretable y explicable.

Fase 2. Riesgos tecnológicos: La empresa auditora pide al hospital diversos insumos para medir los riesgos derivados del uso del sistema *Watson Health*, solicita manuales de instalación, operación y capacitación para conocer el tipo de inteligencia artificial que se está utilizando, los datos que utiliza para emitir diagnósticos y tratamientos y pide una reunión con los ingenieros y médicos responsables del proyecto y de los modelos de algoritmos que utiliza para evaluar las razones por las cuales se tomó la decisión que complicó la situación del paciente hasta perder la vida. También pide información sobre los sesgos algorítmicos que se hayan detectado y los resultados que ha emitido en casos similares dentro del hospital.

Resultados de la revisión de los tipos de inteligencia artificial: De la revisión que hace la empresa auditora al sistema *Watson Health* y a sus manuales de configuración y uso se concluye lo siguiente: a) El sistema de IA se basa en Machine Learning – Deep Learning con características y funciones dirigidas al cómputo cognitivo; b) utiliza algoritmos basados en redes neuronales profundas, así como en sistemas de expertos que interactúan para la toma de decisiones sin intervención directa del ser humano; c) la interfaz del sistema no brinda al ser humano que pudiera ser interpretable para una toma de decisión distinta. En este punto se confirma que se está ante un sistema autónomo que es considerado Agente Artificial Moral Autónomo que toma decisiones sin intervención de los seres humanos y que puede causar riesgos altos según la legislación de la Unión Europea (puede ser sujeto de responsabilidad algorítmica, entendida como la capacidad de decidir y cambiar el entorno del mundo sin la intervención del ser humano).

Resultados de la revisión de los datos.- La empresa auditora se reúne con el equipo responsable de *Watson Health* dentro del hospital para conocer cómo se obtienen los datos del paciente y, posteriormente, cuál es su uso al ser enviados para análisis de *Watson Health*. Se revisan los campos de captura y registro, así como las bases de datos del hospital para conocer si existió algún error al momento

de crear la secuencia de instrucciones en el momento de ser procesadas por la máquina. Se detectan datos estructurados y datos no estructurados y se revisan ambos para ver qué tratamientos da el sistema *Watson Health* y la manera en qué se generan los riesgos de privacidad y transparencia. La selección de algoritmos que se utilizan para obtener información se expone por el responsable del sistema *Watson Health* y se afirma que el usuario es el dueño de los datos ya que nada de su información se utiliza más que para el estudio de su propio caso.

Resultados de la revisión de los sesgos algorítmicos: Se pide al responsable del sistema *Watson Health* que muestre si existen sesgos algorítmicos que se hayan detectado y cuál es la técnica que se utilizó para mitigarlos. Los impactos en los sesgos algorítmicos que busca la empresa auditora tienen que ver con alguna afectación a Derechos Humanos o a un daño que no derive del actuar humano. Se concluye que el sistema *Watson Health* generó sesgos algorítmicos por la manera en cómo se le dotaba de información, siendo su única fuente el Hospital de Oncología de Nueva York.

Resultados de la revisión de la opacidad: Se le cuestiona al responsable de *Watson Health* hasta qué punto sus algoritmos son explicables e interpretables. El proveedor del servicio IBM, manifiesta que, todos sus algoritmos tienen un alto grado de explicabilidad e interpretabilidad y que se someten a revisiones periódicas para evitar la opacidad en sus resultados. La empresa auditora hace pruebas en el sitio de *Watson Health* y concluye que efectivamente no existe un grado de opacidad que traiga consecuencias para los derechos de las personas pero que las fallas de este sistema pueden deberse a la toma de decisiones de seres humanos dentro de la organización IBM ya que se encuentran restringidas las búsquedas sobre bases de datos que seleccionan humanos en Nueva York.

Resultados de la revisión de la interpretabilidad: La empresa auditora solicita conocer la trazabilidad de los algoritmos que utiliza *Watson Health* para llegar a los resultados obtenidos en el último diagnóstico que resultó erróneo. Se realiza una reunión de trabajo con los ingenieros que atienden la cuenta del hospital. Se realizan pruebas en ambiente de producción y se demuestra que el sistema no tiene problemas de funcionamiento ni tampoco genera decisiones que no puedan ser

interpretables o que afecten a las personas en sus Derechos Humanos, no hay una afectación directa al paciente que pueda ser atribuida a la interpretabilidad. Nuevamente se comprueba que las búsquedas se encuentran restringidas por la selección de seres humanos que laboran en IBM, Nueva York.

Resultados de la revisión de los explicabilidad: La empresa auditora solicita conocer el histórico de las decisiones automatizadas que ha tomado el sistema *Watson Health* de todos los pacientes del departamento de oncología. De la revisión hecha se desprende que las circunstancias y el contexto para el tratamiento de datos personales es utilizado y medido de forma adecuada. También se demostró que existen procedimientos matemáticos y estadísticos adecuados para la formación de perfiles de pacientes y que existen riesgos mínimos en el manejo de la información.

Resultados de la revisión de la responsabilidad algorítmica: La empresa auditora concluye que el sistema *Watson Health* no tuvo una falla o un sesgo que lo lleve a una responsabilidad algorítmica y que tanto los datos ingresados como su trazabilidad y resultados son producto de sesgos humanos. El error en la decisión que tuvo se debe a que la información fue filtrada previamente por los médicos adscritos al Memorial Sloan Kettering Cancer Center ya que no permitieron que *Watson Health* utilizara la totalidad de información a la que se comprometió IBM en su contrato, por lo tanto no es un sistema opaco, ni tampoco es un sistema que no permita su interpretabilidad y explicabilidad. Se concluye también que este tipo de errores es de una etapa previa y es fruto de la manipulación del ser humano.

Fase 3: Riesgos éticos. – La empresa de auditoría hace la medición de los riesgos éticos derivados de la toma de decisiones y busca aplicar la Teoría de la responsabilidad de Carlos Llano Cifuentes como un factor indispensable para verificar que la IA sea utilizada con un soporte ético demostrable con indicadores basados en la medición de las decisiones y su justificación. Si bien esta fase no conlleva a la detección de la responsabilidad algorítmica sí contribuye a la argumentación y detección de la responsabilidad civil y penal, ya que una de las responsabilidades de la teoría de Carlos Llano Cifuentes está vinculada a la justificación de las decisiones que toman las personas dentro de las organizaciones.

La empresa auditora pide una reunión con el Consejo General de Administración para explicar la importancia de implementar acciones en su política de integridad basadas en la medición de la responsabilidad moral en la toma de decisiones, pidiendo que cada uno de sus directivos hagan un ejercicio de reflexión respecto a la responsabilidad del consecuente (o responsabilidad de las consecuencias de la acción, que va de la mano con toda la responsabilidad legal de la fase 1), la responsabilidad del antecedente (o responsabilidad de los principios que permea la fase 2) y las responsabilidades del congruente (o responsabilidad de los proyectos anteriores de mi vida) y del trascendente (o de la misión a la que, con mi vida, he sido destinado).

Se les pide que implementen estas reflexiones en su organización, como parte del gobierno corporativo y que no solo que quede en la responsabilidad del consecuente. Las nuevas tecnologías hacen que se olvide los otros tipos de responsabilidad, pero no debe ocurrir ya que esta omisión puede llevar a la empresa a una responsabilidad jurídica. Al analizar el caso *Watson Health* desde un enfoque ético, se dan cuenta que la decisión de la empresa que tomaron de incluirlo como una herramienta de software novedosa resultó ser muy aventurada. El principal error estuvo en que ni el área de *Compliance* ni ninguna otra buscó analizar y justificar su decisión, todos tomaron como suya la decisión del *Memorial Sloan Kettering Cancer Center*, pero nunca evaluaron su viabilidad ni justificaron su uso con argumentos y evidencias de su operación cotidiana. Otra debilidad es que la legislación le da el carácter de software o dispositivo médico que no requiere más allá de una certificación (la cual sí tenía).

La empresa auditora comenta que las cuatro responsabilidades deben ser compartidas por todos los integrantes de la empresa pensando en la búsqueda de un gobierno corporativo con un soporte de integridad superior. Esto a la larga da mejores y mayores resultados. La empresa auditora recomienda la actualización del programa de cumplimiento y del programa de integridad corporativa con base en las cuatro responsabilidades. Además, deben estar conscientes de que su margen de influencia, en muchos casos, es global y por tanto deben ampliar por sí mismas su

radio de responsabilidades y no solo quedarse o atender los efectos inmediatos derivados de sus fines específicos.

Resultados obtenidos de la revisión de la responsabilidad del antecedente: La empresa auditora se reúne con el responsable de cumplimiento del hospital para preguntar si tienen programas de integridad que contemplen el uso de la IA. El responsable del área de cumplimiento del hospital manifiesta que tiene un programa de cumplimiento y un programa de integridad corporativa basado en la aplicación de su código de Ética y su código de conducta, pero no existen evidencias de que esto esté vinculado con el uso de inteligencia artificial y se contemplen acciones de mitigación de las personas que están involucradas en la toma de decisiones. Confirma también que las decisiones directivas no se someten a un procedimiento de verificación de su justificación porque a los directivos no se les cuestiona sus decisiones.

En el caso particular de *Watson Health* se detectó que el hospital dio prioridad a temas relacionados con ventas y mercadotecnia, pero que los directivos no emitieron ninguna justificación Ética ni tampoco pidieron al área de *Compliance* una opinión. El área de *Compliance*, por su parte, aunque debía hacerlo, con base en su manual, se abstiene para evitar detener la puesta en operación de *Watson Health* dejando a un lado la validación y estudio de esta aplicación conforme a los riesgos normativos y éticos.

Resultados obtenidos de la revisión de la responsabilidad del consecuente y del trascendente: La empresa auditora solicita información al área de cumplimiento para conocer sobre las buenas prácticas que existen en materia de gobierno corporativo y de la aplicación de encuestas de clima organizacional para conocer las necesidades de sus empleados y sus intenciones de ir mejorando, desarrollándose y crecer conforme sus planes de vida. No se encuentra ningún indicio de que la empresa utilice estas dos responsabilidades. El informe final se entrega a todos los integrantes del Consejo General de Administración y a todos los directores.

Al final, el hospital decide hacerse cargo del daño y pagar una indemnización a la familia, así como implementar un plan de cumplimiento normativo y de

integridad corporativa con los elementos de las tres dimensiones de responsabilidad del Catálogo de indicadores para medir los riesgos de la IA e incorporarlos en sus matrices de riesgos como parte de aquellos factores que pueden afectar a la organización.

Watson Health suspende operación el 21 de febrero de 2022, entre las razones del fracaso está una mala justificación Ética de las decisiones de sus directivos.²⁶⁰ Esperamos que nuestras reflexiones sean el inicio de una discusión amplia en beneficio de las empresas, gobiernos y cualquier ser humano que utilice IA. Para ello, retomemos lo que Josemaría Escrivá de Balaguer nos compartió:

“Lo que se necesita para conseguir la felicidad, no es una vida cómoda, sino un corazón enamorado. ¡Dios y Audacia!”²⁶¹

²⁶⁰ Cfr. Casey Ross, “Once billed as a revolution in medicine, IBM’s *Watson Health* is sold off in parts” en *STAT*, 21 de enero de 2022, consúltese en <https://www.statnews.com/2022/01/21/ibm-watson-health-sale-equity/> y Sandeep Konam, “Where did IBM go wrong with *Watson Health*?” <https://qz.com/2129025/where-did-ibm-go-wrong-with-watson-health/>, en *Quartz*, 02 de marzo de 2022. [15 de julio de 2022]

²⁶¹ Julián Herranz, *Dios y Audacia*, segunda edición, Ediciones RIALP, Madrid, 2011.

Reflexiones finales

Para concluir esta investigación, y a manera de reflexiones finales, se considera importante retomar algunas ideas que a su vez pueden servir para estudios futuros:

Primera.- La mayor parte de los sistemas de Inteligencia Artificial no producen daños o impactos negativos a los Derechos Humanos de las personas ya que sus algoritmos están controlados de inicio a fin y, por lo general, son un instrumento para que el ser humano realice alguna acción que está dentro de su *expertise* y manipulados como parte de un acto volitivo, ya sea para su puesta en funcionamiento, ya sea para el uso de algún otro ser humano, siendo sus resultados o consecuencias responsabilidad directa de la persona. Sin embargo, existe un tipo de Inteligencia Artificial, que está tomando decisiones, sin intervención ni supervisión del ser humano (o por lo menos, no queda clara su participación), que debe ser analizada e incorporada a los sistemas jurídicos con parámetros de construcción conceptual adicionales a los ya existentes.

Segunda.- Los modelos de IA basados en Machine Learning – Deep Learning y aquellos ubicados en la categoría de sistemas de computación cognitiva, son un tipo de IA que está siendo utilizada para la toma de decisiones y que está siendo regulada en la Unión Europea para evitar daños o impacto a los Derechos Humanos de las personas.

Tercera.- Si un sistema autónomo para la toma de decisiones, por alguna razón, toma una decisión equivocada (o errónea) sin consentimiento o sin la participación del ser humano, el sistema jurídico y sus operadores están responsabilizando a la persona que se encontraba más cercana la tecnología. No hay un juicio de reproche soportado por una norma jurídica (o moral) que permita al juzgador, fiscal, árbitro, que la responsabilidad del sujeto sea categorizada de otra manera. Por lo que

se debe de pensar en la construcción de conceptos adicionales que surjan bajo las nuevas circunstancias que produce este tipo de tecnología.

Cuarta.- El sistema jurídico mexicano, como muchos otros en el mundo occidental, no contempla alguna categoría de responsabilidad jurídica adecuada para resolver este tipo de casos y traslada la responsabilidad de las máquinas a los seres humanos, entre otras razones, porque considera a esta tecnología como un software tradicional, un instrumento o dispositivo que utiliza la voluntad y *expertise* de determinados seres humanos y que, en muchas ocasiones, además fue probado y certificado bajo los estándares de calidad de alguna norma oficial o internacional con características específicas que da el sector de la industria o los servicios que autoriza su uso. Sin embargo, los sistemas autónomos para la toma de decisiones, en los que el ser humano no tiene una interfaz adecuada para entender el origen de la decisión, su trazabilidad y los resultados, deben ser clasificados bajo otra concepción jurídica, tecnológica y ética.

Quinta.- Los operadores de los sistemas de justicia (penal, civil, administrativo, laboral, entre otros) no están dimensionando la problemática, tampoco los legisladores, ni abogados. Por tanto, los sistemas de justicia mantienen sus estándares de argumentación y razonamiento probatorios sin la posibilidad de reducir la responsabilidad del ser humano más cercano a la tecnología (y transferirla a los sistemas autónomos o contemplarlos como una eximente de responsabilidad). El costo social y de afectación a las personas será injusto e inmoral ya que la responsabilidad de los daños o afectación a derechos humanos recaerá en la persona que no tuvo posibilidad de tomar una decisión para controlar al sistema autónomo sin tomar en cuenta que el sistema le restó capacidades de análisis, reacción y solución de la problemática.

Sexta.- Los sistemas de justicia, al generar resoluciones de este tipo, están dejando a las víctimas, ya sea que busquen la reparación del daño por la vía penal, la reparación de daños y perjuicios por la vía civil o algún reclamo por afectación a Derechos Humanos, sin posibilidad de una sentencia moralmente justa, dejando a un lado la responsabilidad de la empresa y del propio sistema autónomo. El ser humano que está más cercano a la tecnología no debe ser el que absorba toda la responsabilidad de la reparación el daño ni tampoco la persona jurídica debe hacerse a un lado o justificar su actuar en certificaciones o manuales de uso que diluyen la responsabilidad jurídica.

Séptima.- Esta problemática debe analizarse desde un enfoque ético (la Filosofía), legal (el Derecho) y tecnológico (la Ingeniería).

Octava.- Si los ingenieros supieran de las consecuencias legales que genera el considerar a un sistema autónomo como un simple software, dispositivo o instrumento (la culpabilidad no disminuye, el grado de participación se mantiene inamovible, no hay juicio de reproche, no hay reparación de daño amplia y ejemplar) y los abogados entendieran que en dicho sistema no participa el ser humano de ninguna manera (o por lo menos, no con la claridad suficiente para delimitar su conducta), seguramente se vería reflejado en la adición de eslabones a las normas jurídicas y a la enseñanza y aplicación del Derecho, en particular, a las categorías de responsabilidad jurídica vigentes.

Novena.- Los avances encontrados, fruto de la discusión y la investigación jurídica, son escasos hasta ahora. Se requiere construir en las universidades, programas específicos sobre IA, Ética y Derecho. Y al exterior, en el sector público, legislación y órganos especializados que realicen una supervisión y control de esta tecnología ya que los sistemas autónomos están presentes en nuestra vida cotidiana y están guiando nuestras decisiones.

Décima.- Se debe fomentar el desarrollo de futuras investigaciones para construir un ecosistema de Inteligencia Artificial (IA) para la protección de los Derechos Humanos de las personas en todo el mundo (o quizás varios sistemas regionales). Esta tarea requiere de varias líneas de investigación y varios expertos trabajando permanentemente en algún centro especializado de creación específica para el logro de este objetivo. La categorización de las reflexiones hechas en esta tesis sobre el uso de la IA y la responsabilidad jurídica pone el piso para que otros investigadores lo hagan con mayor amplitud.

Décima Primera.- La Ética empresarial se debe utilizar como parte de los mitigantes que una para requiere para reducir riesgos. Un ejemplo bastante útil lo encontramos en el modelo de responsabilidad de Carlos Llano Cifuentes y sus cuatro categorías de la responsabilidad. Los abogados e ingenieros debemos sentarnos a revisar esta propuesta teórica.

Décima Segunda.- La construcción de un ecosistema de cuidado social para el uso de IA (o quizás varios ecosistemas regionales) requiere de un conjunto de elementos que los podemos agrupar de manera fundacional en: a) normas morales y jurídicas (dentro y fuera de la empresa); b) agencias reguladoras expertas en IA, unas de corte público, responsables de capacitar, crear políticas, procedimientos y procesos para dirimir controversias, llevar un registro público de algoritmos; supervisar el impacto social y ambiental de esta tecnología; también, agencias que tengan facultades de supervisión y sanción en aquellos casos que se compruebe una afectación a Derechos Humanos; c) empresas certificadoras y auditoras de algoritmos para que midan los posibles riesgos legales de las empresas que utilizan IA y se sumen a los riesgos de la organización; d) Programas y responsables de implementar el *Compliance* al interior de las empresas que crean o utilizan la IA; e) Creación de metodologías para la medición y evaluación del impacto adversos de la IA en los Derechos Humanos; f) Supervisiones periódicas por las agencias reguladoras del gobierno; g) Tribunales especializados en dirimir controversias por temas de responsabilidad algorítmica y afectación a Derechos Humanos por el uso

de este tipo de tecnología; h) Creación de un comité de IA en la estructura de la Organización de las Naciones Unidas (ONU) y sus homólogas regionales, como funciones para casos confirmados de violación de Derechos Humanos por uso de IA, con estructura similar a la de otros comités que existen actualmente; i) actualización de planes de estudio en escuelas, universidades y centros de investigación para incluir los nuevos conceptos y usos de la IA y hacer labores de prevención y de divulgación; j) Participación de los organismos empresariales y cámaras de comercio en la capacitación y divulgación de estos temas; k) utilizar los beneficios de la Ética empresarial y de la participación de filósofos expertos en uso de IA; y, l) el desarrollo de políticas, procesos, procedimientos, sistemas, certificaciones y mecanismos de auditoría o vigilancia al interior y exterior de las empresas que generan IA o de los gobiernos que la estén utilizando para alguna actividad en la que se vea involucrada la población.

Décima Tercera.- Se comprobó la hipótesis planteada en la que se afirma que el uso de algunos modelos de IA que auxilian al ser humano a la toma de decisiones (en este caso, decisiones médicas), están produciendo afectaciones a las personas, algunas de estas afectaciones pueden causar daños cuya responsabilidad está regulada de forma poco clara en la legislación; a la vez puede generar daños o impactos adversos a Derechos Humanos, trasladando la responsabilidad al humano que se encuentre más cercano a la tecnología. No existe en México (ni en la doctrina ni la legislación) alguna solución con la claridad necesaria para categorizar la responsabilidad de las personas, de la empresa o de las máquinas.

Décima Cuarta.- Las empresas de IA, como cualquier empresa, deben ser entendidas como una comunidad de personas que, a través de sus directivos, toma decisiones. Estas decisiones a veces pueden ser certeras y exitosas y a veces pueden ser erróneas y dañinas (generando daños y/o impactos adversos a los Derechos Humanos). Cuando una o varias empresas, a través de sus directivos, deciden utilizar o poner al servicio de sus clientes o usuarios un sistema de IA, deben establecer mitigantes de tres tipos: a) Mitigantes jurídicos, que deben ser

tomados del sistema constitucional o convencional, para prevenir riesgos relacionados con la legislación del sector en el que participa la empresa; b) Mitigantes tecnológicos, basados en el tipo de Inteligencia Artificial que se utiliza y tomando en cuenta las directrices o guías que han emitido las asociaciones de ingenieros expertas en estos temas y sumarlas a los controles jurídicos, para prevenir riesgos por el uso de las nuevas tecnologías; y c) Mitigantes éticos, para mitigar riesgos derivados del comportamiento humano y del uso de la tecnología que permitan crear una cultura basada en la integridad y en el fomento de valores, principios y reglas de comportamiento de las personas, organizaciones y, en algún momento, de los sistemas autónomos.

Décima Quinta.- Existen un conjunto de conceptos que deben ser incorporados a al análisis de abogados, filósofos e ingenieros. En particular: Agencia Moral Artificial Autónoma, justicia algorítmica, esos algorítmicos y responsabilidad algorítmica. Así como evolucionó la responsabilidad de las personas jurídicas, así deberá añadirse el concepto de responsabilidad algorítmica y el uso de auditorías algorítmicas para la protección de la empresa y el ser humano.

Décima Sexta.- Creemos que el Instituto de Humanidades y el Instituto Panamericano de Alta Dirección, de la Universidad Panamericana, deben abrir líneas de investigación y participar en el desarrollo de estos temas, articulando cuerpos de investigación y docencia transdisciplinarios para formar a la alta dirección de las empresas y a su equipo de administración de riesgos o *Compliance* en este tipo de fenómenos que surgen por el uso de la Inteligencia Artificial. El caso Watson, se concluye, es un ejemplo de falta de controles en la organización -legales y éticos- que pudo prevenirse con la medición de riesgos legales, tecnológicos y, sobre todo, éticos, tal como alude Carlos Llano Cifuentes.

Glosario

Las definiciones contenidas en este glosario fueron seleccionadas del texto de este trabajo de investigación tomando en cuenta los términos más utilizados por la comunidad que estudia la Inteligencia Artificial, la Ética y el Derecho. También contiene conceptos propios que fueron contruidos como resultado de la investigación.

Glosario

A

Algoritmo. – Es un concepto multívoco, consúltase la tabla de la página 36.

Algoritmos de caja negra. - Se dice que es de caja negra el algoritmo del que no es posible dar una interpretación de lenguaje natural y referida a los valores de las predicciones y del porqué el algoritmo toma determinada decisión.

API. - Significa interfaz de programación de aplicaciones (*Application Program Interface*).

Aprendizaje profundo.- El aprendizaje profundo es el subcampo del aprendizaje automático que diseña y evalúa arquitecturas y algoritmos de entrenamiento para modelos de redes neuronales modernas. Una red neuronal profunda es una red que tiene múltiples capas de neuronas. (Kelleher)

Aprendizaje reforzado.- El objetivo del aprendizaje por refuerzo es permitir que un agente aprenda una política sobre cómo debe actuar en un entorno determinado. Una política es una función que mapea desde las observaciones actuales de un agente sobre su entorno y su propio estado interno hasta una acción. Por lo general, se usa para tareas de control en línea, como el control de robots y juegos. (Kelleher)

Aprendizaje supervisado.- Una forma de aprendizaje automático donde el objetivo es aprender una función que mapea a partir de un conjunto de atributos de entrada. (Kelleher)

Atributo: Característica o rasgo de una entidad que describe la entidad; por ejemplo, el número de teléfono de un empleado es uno de los atributos de empleado. (IBM)

Agente Artificial Moral Autónomo (AAMA).- Es una creación de la Ética para la IA. Busca o permite dividir la responsabilidad que le corresponde a la máquina (responsabilidad algorítmica) de la responsabilidad que les corresponde a los seres humanos (responsabilidad objetiva, subjetiva, contractual, extracontractual, etcétera). Esta nueva concepción del agente y de la agencia artificial moral autónoma, permite construir y dar soporte a la categoría de responsabilidad algorítmica. Por lo menos en el mundo occidental, de 2018 a la fecha se ha venido dando un movimiento de muchos científicos, en su mayoría formados en Filosofía y Ética, que ha cobrado gran relevancia para el Derecho y la forma en la que se debe regular la reparación del daño por una afectación a Derechos Humanos cuyo origen encontramos en modelos automatizados para la toma de decisiones de los sistemas inteligentes basados en algoritmos como las Redes Neuronales Profundas. En este tipo de tecnología, la máquina de manera independiente, con los datos que se le proporciona o que obtiene del ambiente, va proponiendo soluciones a temas específicos. Es decir, no existe la voluntad humana presente en la construcción de

soluciones. La teoría de la autonomía de la voluntad y el libre albedrío, la capacidad de razonar, conocer y efectuar el acto con consciencia e inteligencia, a través de la voluntad del hombre, parecieran ser poco útiles para calcular la responsabilidad algorítmica. No hay un ser humano que consciente y voluntariamente haga una programación de la máquina que emite el resultado, ni un ser humano que esté guiándola en la búsqueda de información o el filtrado de datos. En el mejor de los casos, quizás encontremos a una persona que ingresa datos -el médico, su auxiliar o algún empleado que administra los equipos del hospital o que busca datos para que sean incorporados o incluidos a la base de datos o de información y éstos a su vez son analizados y procesados conforme la base de datos o de la información el campo de la medicina o de la profesión que se tenga autorizado para su uso. Pero no existen aún en México y en América Latina normas, instituciones y personas - físicas o jurídicas- que supervisen y auditen el uso de estos datos y de los algoritmos de inteligencia artificial que se utilizan para la toma de decisiones. Para determinar la responsabilidad algorítmica de un AAMA, se debe evaluar el tipo de algoritmo que se está utilizando, los datos, el modelo de IA, el tipo de algoritmo, el modelo de IA, si es opaco o no, si hubo o no sesgos algorítmicos, la interpretación y explicabilidad, entre otros.

B

Base de Datos (Dataset). - La información almacenada en varios archivos interrelacionados.

C

Computación cognitiva. - Es el término que se usa para describir sistemas de inteligencia artificial que tienen como objeto simular el pensamiento humano. La cognición humana implica el análisis en tiempo real del entorno, el contexto y la intención, entre muchas otras variables que conforman la capacidad de una persona para razonar y resolver problemas.

Conjunto de datos.- Una colección de instancias con cada instancia descrita en términos de un conjunto de características. En su forma más básica, un conjunto de datos se organiza en una matriz $n \times m$, donde n es el número de instancias (filas) y m es el número de características (columnas). (Kelleher)

Cuarta revolución industrial. - Comenzó a partir de este siglo y se basa en la revolución digital. Se caracteriza por un internet más ubicuo y móvil, por sensores más pequeños y potentes que son cada vez más baratos, y por la inteligencia artificial y el aprendizaje de la máquina

D

Datos ciegos: Un conjunto de documentos anotados con los datos de campo, como pares de preguntas y respuestas, anotación semántica y decisiones de pasaje. Los datos ciegos nunca los lanzan ni los ven los desarrolladores y se utilizan para probar el sistema periódicamente para evaluar el rendimiento en datos no mostrados. La prueba en datos ciegos impide que se corrompa la exactitud al sobre ajustar a conjuntos o anotaciones de preguntas conocidas. Los resultados notificados solo deberían provenir de las pruebas que se ejecutan en los datos ciegos. (IBM)

Datos de entrenamiento: Un conjunto de documentos anotados que pueden utilizarse para entrenar modelos de aprendizaje automático. (IBM)

Datos de prueba: Un conjunto de documentos anotados que pueden utilizarse para evaluar las métricas de sistema después de la ingesta y del entrenamiento. (IBM)

Deep Learning o Aprendizaje Profundo.- El aprendizaje profundo permite que los modelos computacionales que están compuestos por múltiples capas de procesamiento aprendan representaciones de datos con múltiples niveles de abstracción. Estos métodos han mejorado drásticamente el estado de la técnica en reconocimiento de voz, reconocimiento de objetos visuales, detección de objetos y muchos otros dominios, como el descubrimiento de fármacos y la genómica.

Diccionario: Una recopilación de palabras que pueden utilizarse para preanotar documentos. Se creará una anotación nueva para cada palabra en el texto del documento que coincida con un término del diccionario. Un modelo de aprendizaje automático se puede configurar con uno o varios diccionarios independientes, que son normalmente específicos del dominio, como un diccionario para productos farmacéuticos y un diccionario para la gestión de la riqueza. (IBM)

E

Ecosistema de cuidado social para el uso de IA que proteja los Derechos Humanos.- Se requiere un ecosistema de cuidado social para el uso de IA que proteja los Derechos Humanos o quizás varios ecosistemas regionales de protección de Derechos Humanos. Este nuevo paradigma se debe construir transdisciplinariamente. Y como propuesta para la construcción de un ecosistema de cuidado social para el uso de la IA consideramos prudente enunciar los ingredientes básicos para su funcionamiento: a) normas, ya sea morales o jurídicas (de *soft law* o *hard law*); b) agencias reguladoras del gobierno expertas en IA, unas de corte público, responsables de educar, crear políticas, procedimientos, procesos y normas; dirimir controversias, llevar un registro público de algoritmos; certificar y auditar; supervisar el impacto social y sancionar aquella IA que afecte Derechos Humanos; c) Auditores privados, para realizar auditorías de Derechos Humanos a empresas y gobiernos, así como para medir riesgos en esta materia; d) Áreas de

Compliance que se constituyan al interior de las grandes empresas que crean o utilizan IA para la medición de riesgos empresariales (con base en los postulados de Carlos Llano Cifuentes); e) Creación de metodologías para la medición y evaluación del impacto de la IA en los Derechos Humanos; f) Supervisiones periódicas por las agencias reguladoras; g) Tribunales especializados en dirimir controversias que lleguen a sede judicial por temas de responsabilidad algorítmica y afectación a Derechos Humanos; h) Actualización de normas internas; i) Creación de un comité de IA en la estructura de la ONU y sus homólogas regionales, con funciones para casos confirmados de violación de Derechos Humanos por uso de IA, con estructura similar a la de otros comités que existen actualmente ; j) actualización de planes de estudio en escuelas, universidades y centros de investigación para incluir los nuevos conceptos y usos de la IA y hacer labores de prevención y de divulgación; k) Participación de los organismos empresariales y cámaras de comercio en la capacitación y divulgación de estos temas; y, l) utilizar los beneficios de la Ética en la IA.

Entidad: Mención anotada por un tipo de entidad. Persona, objeto o concepto sobre el que se almacena información. Conjunto de detalles que se mantienen sobre un objeto del mundo real como una persona, una ubicación o una cuenta bancaria. Una entidad es un tipo de elemento. (IBM)

Entrenar: Proceso de configurar una instancia de Watson con componentes que permiten que el sistema funcione en un dominio determinado (por ejemplo: contenido del corpus, datos de entrenamiento que generan modelos de aprendizaje automático, algoritmos mediante programación u otros componentes de datos de campo) y luego de realizar mejoras y actualizaciones en estos componentes basándose en el análisis de exactitud. (IBM)

Errores cognitivos (sesgos cognitivos). - Ver definiciones de la tabla de la página 150.

Ética basada en el consecuencialismo. - Este enfoque sostiene que las propiedades normativas de un acto dependen únicamente de las consecuencias de ese acto. Por lo tanto, un acto se considera moralmente correcto puede determinarse examinando las consecuencias de ese acto: ya sea del acto en sí (utilitarismo del acto, asociado con Jeremy Bentham) o de la existencia de una regla general que requiere actos de ese tipo (utilitarismo de las reglas, a menudo asociado con John Stuart Mill). Esto da lugar a la pregunta de cómo se evalúan las consecuencias. Tanto Bentham como Mill creían que debía prevalecer en las decisiones humanas la regla de “la mayor felicidad del mayor número”, aunque Mill adoptó una visión más refinada de lo que debería considerarse felicidad. Hay una serie de críticas asociadas con esta noción y han sugerido muchas variedades. Hay críticas basadas en la imposibilidad de calcular todas las consecuencias de cada acto para cada persona. También se dice que el consecuencialismo no pretende ser un procedimiento de decisión, sino un criterio para juzgar las acciones. A pesar de

la dificultad de determinar todas las consecuencias reales de un acto (por no hablar de una regla), generalmente hay buenas razones para creer que una acción (o regla) aumentará o disminuirá la utilidad general, y tales razones deberían guiar las elecciones de un consecuencialista. Según el último punto de vista, las acciones se juzgan no en función de las consecuencias reales, sino de las consecuencias esperadas, dada la probabilidad de los diversos futuros posibles. Dada nuestra noción de un agente débilmente ético, el agente mismo recibirá una función de utilidad y elegirá acciones que intenten maximizarla. La elección de la función de utilidad y la forma en que se calculan las consecuencias serán responsabilidad del diseñador.

Ética deontológica. - El elemento clave de la Ética deontológica es que el valor moral de una acción se juzga por su conformidad con un conjunto de reglas, independientemente de sus consecuencias. Un ejemplo es la Filosofía Ética de Kant. Por ejemplo, en el corazón de la obra de Kant están el segundo imperativo categórico, el concepto de dignidad. Aquí tenemos un principio, el no tratar al ser humano como un objeto sino como un fin en sí mismo. Por tanto, uno debe actuar solo de acuerdo con ese precepto que desearía que se convierta en una ley universal, de modo que las reglas mismas se basen únicamente en la razón. Otra forma de actuación la ofrece la Teoría de la Justicia de Rawls, en el que las normas corresponden a principios aceptables bajo un contrato social adecuadamente descrito. Las reglas a seguir serán elegidas por el diseñador, y el agente moral será un mero seguidor de reglas. Así, el agente moral encarnará sólo una parte muy poco sofisticada de la deontología: cualquier sofisticación reside en el diseñador que desarrolla el conjunto de reglas que seguirá el agente. Los problemas de la Ética deontológica incluyen la posibilidad de conflictos normativos y el hecho de que obedecer una regla puede tener consecuencias claramente indeseables. Son muchas las situaciones en las que se puede considerar incorrecto obedecer una ley del país, y no es difícil imaginar situaciones en las que hay argumentos de que sería incorrecto obedecer también una ley moral. Parte de esto puede manejarse mediante excepciones (que pueden verse como modificaciones que legitiman la violación de la regla general en ciertas circunstancias prescritas) a las reglas. Sin embargo, prever todas las excepciones es tan imposible como prever todas las consecuencias de una acción. Por esta razón, las leyes a menudo se redactan con cierto grado de vaguedad. Así los jueces pueden interpretar y decidir casos de hechos particulares.

Ética de la Virtud. - La Ética de la virtud es el más antiguo de los tres enfoques y se remonta a Platón y Aristóteles. La idea básica aquí es que las acciones moralmente buenas ejemplificarán virtudes y las acciones moralmente malas ejemplificarán vicios. La Ética de la virtud tradicional se basa en la noción de “*eudaimonia*”, generalmente traducida como felicidad o florecimiento. La idea aquí es que las virtudes son rasgos que contribuyen a, o son un componente de la *eudaimonia*. Las alternativas son; Ética de la virtud basada en agentes, que

entiende lo correcto en términos de buenas motivaciones y lo incorrecto en términos de tener motivos malos (o insuficientemente buenos), la Ética de la virtud centrada en el objeto, que sostiene que ya tenemos una idea aceptable de qué rasgos son virtudes y lo que implican, y la Ética platónica de la virtud, inspirada en la discusión de las virtudes en los diálogos de Platón. Existe una amplia gama de Ética de la virtud, pero todos ellos tienen en común la idea de que una característica importante de la Ética de la virtud es que reconoce diversos tipos de razones morales para la acción, y tiene algún método (correspondiente a la *phronesis*: prudencia o sabiduría práctica) por considerar estas cosas al decidir cómo actuar.

Ética algorítmica.- Se pregunta cómo la IA, aprendizaje máquina y robots interpretan los datos. (Aníbal Monasterio)

Ética de la IA.- Prioriza consideraciones para la creación de sistemas de IA, promoviendo estándares o difundiendo políticas globales, regionales o nacionales para su uso. (Aníbal Monasterio)

Ética de las máquinas.- Busca construir sistemas inteligentes que tomen decisiones autónomas y se pregunta cómo codificar la moral en las máquinas; hay también lo que se llama “IA moral”, dedicada a conseguir la cooperación máquina-ser humano para resolver problemas complejos con el uso de la moral; la IA es vista como un mecanismo para la mejora de los seres humanos. (Aníbal Monasterio)

Exactitud: Una medida de la corrección de anotaciones producidas por un modelo de aprendizaje automático. (IBM)

Exhaustividad: Una medida que especifica el porcentaje de resultados relevantes devueltos, de todos los resultados relevantes disponibles. La exhaustividad, que es una medida de sensibilidad, está determinada por el número de resultados positivos correctos dividido por el número de resultados positivos que deberían haberse devuelto. La exactitud se mide mejor mediante la precisión y la exhaustividad. (IBM)

Explicabilidad: El Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial creado por la Comisión Europea, comenta que es un “principio ético crucial” para conseguir que los usuarios confíen en los sistemas de IA y para mantener dicha confianza. La finalidad de los sistemas de IA y que las decisiones deben poder explicarse —en la medida de lo posible— a las partes que se vean afectadas por ellas de manera directa o indirecta.

F

Falso negativo: Una respuesta o una anotación que es correcta, pero que se ha predicho que es incorrecta.

Falso positivo: Una respuesta o anotación que es incorrecta, pero que se ha predicho que es correcta.

H

Healthcare technology.- Es definida por la Organización Mundial de la Salud (OMS) como “la aplicación de conocimientos y habilidades organizados en forma de dispositivos, procedimientos, sistemas, medicamentos y vacunas desarrollados para salvar un problema de salud y mejorar la calidad de vida. (OMS)

Heurística. - La heurística, según el Diccionario de la Lengua Española es la “Técnica de la indagación y el descubrimiento”. Gabriel Zaid nos comenta que Herbert A. Simon propuso en 1957 el desarrollo de programas de computación heurística para que las máquinas tomen decisiones, a partir de criterios mecánicos.

I

Inteligencia Artificial. - Es un concepto multívoco, consúltase la tabla de la página 34.

Interpretabilidad: El principio de interpretabilidad parte de la siguiente hipótesis: Si el sistema puede explicar su razonamiento, entonces los seres humanos pueden verificar si ese razonamiento es correcto, completo, adecuado con respecto otros criterios, normas, reglas, derechos. Visto esto como un principio que se debe intentar cumplir al usar IA y, principalmente algoritmos de IA basada en Machine Learning – Deep Learning , la interpretabilidad es la capacidad para explicar o presentar en términos comprensibles algo a un ser humano. Interpretar en el ambiente de la IA significa explicar o presentar en términos comprensibles. En el contexto de los sistemas de IA basada en Machine Learning – Deep Learning , definimos la interpretabilidad como la capacidad de explicar o presentar en términos comprensibles a un humano.

J

Justicia algorítmica. - es una rama nueva del conocimiento jurídico que explora las preocupaciones técnicas, legales y Éticas que existen en la toma de decisiones con base en algoritmos, en particular cuestiones de sesgos algorítmicos, que pueden afectar Derechos Humanos. (José Luis Hernández)

L

Lenguajes de programación. - son las herramientas con las que el ser humano instruye a las máquinas lo que deben hacer. Sin estos, no hubiera sido posible el funcionamiento y evolución de la tecnología que hoy tenemos.

M

Machine Learning o Aprendizaje Automático. - Es un concepto multívoco, consúltese la tabla de la página 96.

Medicina Basada en Evidencia. - La Medicina Basada en la Evidencia "...precisa la integración de los mejores indicios obtenidos a través de la investigación, nuestros conocimientos teóricos y prácticos y los valores y circunstancias específicos de cada paciente. • Con la expresión *mejores indicios obtenidos a través de la investigación* nos referimos a las investigaciones de interés clínico, que en ocasiones proceden de las ciencias básicas de la medicina, pero sobre todo de las investigaciones clínicas centradas en el paciente para determinar la exactitud y precisión de las pruebas diagnósticas (incluida la exploración clínica), la potencia de los marcadores pronósticos y la eficacia y seguridad de las estrategias terapéuticas, rehabilitadoras y preventivas. • Por *conocimientos teóricos y prácticos* entendemos la capacidad para utilizar nuestras habilidades clínicas y la experiencia adquirida para identificar rápidamente el estado de salud y el diagnóstico específicos de cada paciente, los riesgos y los beneficios derivados de las posibles intervenciones/ exposiciones/ pruebas diagnósticas en cada caso concreto y sus valores y expectativas personales. (Straus, Sharon E.)

Modelo.- En el aprendizaje automático, un modelo es un programa informático que codifica los patrones que el algoritmo de aprendizaje automático ha extraído de un conjunto de datos. Hay muchos tipos diferentes de modelos de aprendizaje automático; sin embargo, el aprendizaje profundo se centra en la creación de modelos de redes neuronales con múltiples capas de neuronas ocultas. Un modelo se crea (o entrena) ejecutando un algoritmo de aprendizaje automático en un conjunto de datos. Una vez que se ha entrenado el modelo, se puede utilizar para analizar nuevas instancias; el término inferencia a veces se usa para describir el proceso de analizar una nueva instancia usando un modelo entrenado. En el contexto del aprendizaje automático, los términos modelo y función a menudo se usan indistintamente: un modelo es una instanciación de una función como un programa de computadora. (Kelleher)

Modelo matemático, en su forma simple y para ilustrar nuestra investigación, es "una ecuación que describe cómo una o más variables de entrada se relacionan con una variable de salida. De esta forma, un modelo matemático es lo mismo que una función: un mapeo de entradas a salidas. (Kelleher)

N

Neurona.- En el contexto del aprendizaje profundo (a diferencia de la ciencia del cerebro), una neurona es un algoritmo de procesamiento de información simple que toma una cantidad de valores numéricos como entrada y asigna estos valores a una activación de salida alta o baja. Este mapeo generalmente se implementa multiplicando primero cada valor de entrada por un peso, luego sumando los resultados de estas multiplicaciones y finalmente pasando los resultados de la suma ponderada a través de una función de activación. (Kelleher)

O

Orden público tecnológico. - Es el conjunto de medidas tomadas por los poderes públicos, tendientes a organizar las relaciones de los usuarios de los bienes y servicios tecnológicos con los responsables de su desarrollo o producción a fin de evitar un aprovechamiento indebido por parte de éstos debido a las dificultades generadas por la superioridad de conocimientos y el acceso a la tecnología

Orden público no tecnológico. - Es el conjunto de condiciones fundamentales de vida social instituidas en una comunidad jurídica, las cuales, por afectar centralmente a la organización de ésta, no pueden ser alteradas por la voluntad de los individuos”. En palabras de Granero, el orden público tecnológico “establece en la esfera de la autonomía de la voluntad privada un régimen de limitaciones relativas al contenido de ciertas relaciones jurídicas por razones de bien común.

P

Patrones oscuros. - Término que se utiliza para describir interfaces en línea en sitios web y aplicaciones diseñadas para manipular intencionalmente a los usuarios para que tomen decisiones que de otro modo no tomarían en circunstancias normales.

Precisión: Una medida que especifica la proporción de resultados que son relevantes. La precisión, que es un valor predictivo positivo, se determina mediante el número de resultados positivos correctos dividido por el número de todos los resultados positivos. La exactitud se mide mejor mediante la precisión y la exhaustividad. (IBM)

Principio de transparencia. - Exige que toda información y comunicación relativa al tratamiento de dichos datos sea accesible y fácil de entender, y que se utilice un lenguaje sencillo y claro. Dicho principio se refiere en particular a la información de los interesados sobre la identidad del responsable del tratamiento y los fines de este

y a la información añadida para garantizar un tratamiento transparente con respecto a las personas físicas afectadas y a su derecho a obtener confirmación y comunicación de los datos personales que les conciernan

R

Responsabilidad algorítmica o responsabilidad por la toma de decisiones basada en algoritmos.- Es la que se produce por un Agente Autónomo Moral Artificial (AAMA), que toma decisiones a través de sus propios algoritmos, sin la participación de la voluntad o decisión de un ser humano. El Parlamento Europeo, el 16 de febrero de 2017, creó la categoría jurídica de Persona Electrónica o Persona Algorítmica. Cuando la máquina toma decisiones de manera automática o autónoma con el solo procesamiento de los datos que obtiene de las bases de datos o del entorno, estamos ante una máquina que puede ser sujeto de responsabilidad; cuando la máquina no es capaz de aprender por sí sola y requiere de un ser humano, de su voluntad y consentimiento, estamos ante una categoría ajena a la responsabilidad algorítmica. La responsabilidad algorítmica es un concepto estudiado por la justicia algorítmica. Este es un tipo de responsabilidad que no se tenía contemplada ni por los juristas del siglo XIX ni por los juristas del siglo XX, o por lo menos no con el panorama que estamos viviendo a partir del descubrimiento de sistemas inteligentes que generan soluciones automatizadas sin la intervención de la voluntad y el conocimiento humano y es la llamada responsabilidad algorítmica o responsabilidad por la toma de decisiones basada en algoritmos. La teoría de la autonomía de la voluntad y el libre albedrío, la capacidad de razonar, conocer y efectuar el acto con consciencia e inteligencia, a través de la voluntad del hombre, parecieran ser poco útiles para calcular la responsabilidad algorítmica. La responsabilidad algorítmica, a pesar de no ser norma vigente, tiene componentes que requieren de su positivización o aplicación adicional a la norma jurídica para evitar daños a la empresa por la no mitigación de riesgos tecnológicos. (Ver “justicia algorítmica”)

Rendimiento: Medida de un sistema de Watson en términos de exactitud, precisión y exhaustividad, por ejemplo, al responder preguntas, descubrir relaciones o anotar texto. (IBM)

Retropropagación o Backpropagation.- es un algoritmo utilizado para entrenar una red neuronal con capas ocultas de neuronas. Durante el entrenamiento, los pesos en una red se actualizan iterativamente para reducir el error de la red. Para actualizar los pesos de los enlaces que llegan a una neurona específica en una red, primero es necesario calcular una estimación de la contribución de la salida de esa neurona al error general de la red. El algoritmo de retropropagación es una solución para calcular estas estimaciones para cada neurona de la red. Una vez que se han

calculado estas estimaciones de errores para cada neurona, los pesos de las neuronas se pueden actualizar utilizando un algoritmo de optimización como el descenso de gradiente. La retropropagación funciona en dos fases: un paso hacia adelante y un paso hacia atrás. En el paso hacia adelante, se presenta un ejemplo a la red y el error general de la red se calcula en la capa de salida de la red comparando la salida de la red con la salida esperada para el ejemplo especificado en el conjunto de datos. En el paso hacia atrás, el error de la red se comparte a través de la red y cada neurona recibe una parte de la culpa del error en proporción a la sensibilidad del error a los cambios en la salida de esa neurona. El proceso de compartir los errores a través de la red se conoce como retropropagación de los errores y aquí es donde el algoritmo recibe su nombre. (Kelleher)

Red neuronal convolucional.- Una red neuronal convolucional es una red que tiene al menos una capa convolucional. Una capa de convolución está compuesta por un conjunto de neuronas que comparten el mismo conjunto de pesos y cuyos campos receptivos combinados cubren una entrada completa. La unión de las salidas de dicho conjunto de neuronas se conoce como mapa de características. En muchas redes neuronales convolucionales, los mapas de características se pasan a través de una capa de activación de ReLU y luego una capa de agrupación. (Kelleher)

S

Sistemas Expertos Decisionales.- Ver definición de la página 51.

Sesgo. - En la definición gramatical nos referimos a un error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras. Podemos decir que existen sesgos generados por las personas, sesgos generados por las máquinas/sistemas y sesgos de la sociedad.

Sesgo algorítmico. - Aquel funcionamiento del sistema de IA basada en Machine Learning – Deep Learning que no genera los resultados esperados afectando con ello, directa o indirectamente, a personas o a ciertos subgrupos sociales, abarcando disparidades junto con atributos que son categorías protegidas en los tratados y convenciones internacionales de Derechos Humanos, como la clase socioeconómica o el nivel educativo, el sexo, la raza, la orientación sexual, la edad, alguna discapacidad y otros atributos relacionados con la dignidad de la persona y sus Derechos Humanos (la vida, la libertad, la propiedad, etcétera). (José Luis Hernández)

T

Teoría de una Moral de Responsabilidad Empresarial.- Según Carlos Llano Cifuentes existen cuatro campos de actuación de las personas: 1) la responsabilidad del consecuente (o responsabilidad de las consecuencias de la acción); 2) la responsabilidad del antecedente (o responsabilidad de los principios); 3) la responsabilidad congruente (o responsabilidad de los proyectos anteriores de mi vida); y 4) la responsabilidad del trascendente (o de la misión a la que, con mi vida, he sido destinado). La importancia de utilizar esta teoría como ingrediente adicional a los controles de autorregulación o de integridad corporativa que las empresas tienen y como un complemento sus programas de *Compliance*, les permitirá demostrar a la autoridad que la empresa ha hecho todo lo que está en sus manos para evitar una afectación derivada de sistemas de IA y evitará así, una posible sanción jurídica o reputacional.

.

Toma de decisiones algorítmica. - se refiere a decisiones tomadas utilizando un sistema de IA basada en Machine Learning – Deep Learning. Estas decisiones pueden automatizarse completamente o utilizarse para informar a los tomadores de decisiones humanos.

W

Well-Being Impact Assessment.- Es el proceso mediante el cual uno identifica los impactos potenciales en el bienestar a través de la participación de las partes interesadas, mide dichos impactos a través de la recopilación y el análisis de datos, y modifica un sistema de IA con base en los conocimientos derivados de la participación de las partes interesadas y el análisis de datos. Este proceso es un proceso iterativo que debe repetirse a lo largo del ciclo de vida del sistema y se compone de cinco actividades: a) Análisis interno y participación de las partes interesadas / usuarios; b) Creación del tablero de indicadores de bienestar; c) Creación de un proceso de recopilación de datos y la recopilación de datos; d) Análisis de datos y uso de dichos datos para mejorar el sistema; y e) Iteración.

Índice de tablas, gráficas y diagramas

Índice de tablas y gráficas

Tabla de definiciones sobre “Inteligencia Artificial” o “Sistemas de Inteligencia Artificial”	33
The AI Stack	34
Definiciones de algoritmo	36
Prueba de Touring	38
Características del lenguaje utilizado en Logic Theory Machine (LT)	39
Diagrama de flujo del LT	40
Código binario o código máquina	42
Ejecución y fases de un programa	43
Lenguajes de programación	45
Programación orientada a objetos	47
Common expert systems / rules engines	52
Primeros Sistemas Expertos Decisionales	55
Ventajas y desventajas de los Sistemas Expertos Decisionales	63
Ubicación de los Sistemas Expertos Decisionales dentro del campo de la Inteligencia Artificial	64
Sistemas Expertos Decisionales: Principales metodologías y aplicaciones para la salud	72
Principales temas de los Sistemas Expertos Decisionales en los servicios de salud	75
Significance of each principle and CDSS design considerations	78
Arquitectura de INTEGRA	79
DeepMind (evolución de sus algoritmos)	86
MuZero: Dominando Go, ajedrez, shogi y Atari sin reglas	87
Tiempos de aprendizaje de los algoritmos de DeepMind	87
Definiciones sobre “ <i>Machine Learning</i> ” o “Aprendizaje Automático”	95
Descripción del procesamiento de datos en redes neuronales	111
Tipos de algoritmos basados en redes neuronales profundas	116
Selección de algoritmos según el problema a resolver (reglas básicas)	120

Ejemplos de algoritmos y su relación con el ML	124
Técnicas de ML utilizadas para la predicción de enfermedades	128
<i>Machine Learning vs Deep Learning</i>	133
Tipos de errores cognitivos (sesgos cognitivos)	149
Problemas que causan sesgos y acciones de mitigación recomendadas	154
Fortalezas, beneficios y desafíos de los sistemas RNP XAI actuales	169
Spectrum of Disclosure	181
Proyectos legislativos en el ayuntamiento de Nueva York que regulan los algoritmos o la IA	212
Herramientas para la realización de una auditoría algorítmica con EBR	229
¿Qué es Watson? ¿Cómo funciona?	241
Portal Watson, IBM (junio, 2020)	241
Funciones clave de Watson	243
Funciones de Watson en la nube de IBM -pantalla principal- (abril, 2021) Parte 1	244
Funciones de Watson en la nube de IBM -pantalla principal- (abril, 2021) Parte 2	245
Resultados del estudio del “cuadrante mágico” de Gartner	272
Infraestructura para la nube híbrida (abril 2021)	285
Esperanza de vida de los seres humanos en 1800, 1950 y 2015	295
Principales pandemias en el mundo	296
Avances científicos y tecnológicos que han salvado vidas	299
Formulación de una pregunta clínica	304
Niveles de organización de la evidencia: estructura 6S	305
Tendencias digitales y salud	312
Conference on Connected Health: Applications, Systems and Engineering Technologie	315
Fusiones y adquisiciones de empresas de tecnología para la salud digital	317
Factores existentes en los sistemas de salud	319
Empresas que están compitiendo y generando cambios en la industria de la salud	321
Asociaciones o alianzas para fortalecer <i>Watson Health</i>	332

Digital Health Hype Cycle 2020	345
Ocho principios rectores de la transformación digital del sector de la salud	353
Fuente y destino del daño y excluyentes de responsabilidad	398
Elementos del delito	400
Integración de los elementos positivos y negativos del delito	400
Teoría del garantismo penal y sus diez principios	402
Mala praxis médica (categorías)	415
Factores que generan errores médicos	416
Problemas de malas prácticas médicas y consecuencias	421
Diferencias entre moral humana y moral artificial	437
Cuadrícula de Blake	476
Lista de preguntas básicas para inculcar una cultura de respeto a las dimensiones legal, tecnológica y ética	491
Catálogo de indicadores básicos para la medición de riesgos por el uso de IA	491

Tablas y gráficas por anexo (capítulos I a III)

Capítulo I. Definiciones

Anexo 1

Dartmouth Summer Research Project on Artificial Intelligence 626

Anexo 2

Some Studies in Machine Learning Using the Game of Checkers Arthur
L. Samuel (1959) 629

Anexo 3

A Theory of Transfer Learning with Applications to Active Learning Liu Yang,
Steve Hanneke y Jaime Carbonel 630

Anexo 4

The perceptron: a Probabilistic Model for Information Storage and Organization
in the Brain 631

Anexo 5

Electronic "Brain" Teaches Itself 633

Anexo 6

Van Veen, F. y Leijnen, S. (2019). "The Neural Network Zoo". 634

Capítulo II. Problemas de funcionamiento de la Inteligencia Artificial pro-ML – DL y su regulación

Anexo 1

"Principles for accountable algorithms and a social impact statement
for algorithms 642

Anexo 2

Principios para la administración responsable de un IA confiable de la Recomendación del Consejo de Inteligencia Artificial	646

Anexo 3

Ejemplo de política de privacidad (marzo 2021)	648

Capítulo III. IBM Watson y el caso *Watson Health*

Anexo 1

Editor de reglas	650

Anexo 2

Adición de diccionarios a un espacio de trabajo	652

Anexo 3

Glosario IBM	654

Anexo 4

Historia de IBM	661

Anexo 5

Definición de la palabra cáncer en el <i>Theatrum praxeos medicae</i>	663

Fuentes

Fuentes

1.- Libros y revistas especializadas

Academia de la Lengua Española, *Diccionario de la lengua española*, Vigésima Primera Edición, Madrid, Espasa-Calpe, 1992.

Academia Nacional de Ciencias de Buenos Aires, Centro de Estudios Tecnológicos Inteligentes, “Ciclo de conferencias sobre “Sistemas Inteligentes y Ética”. Consúltase las sesiones en: <https://www.youtube.com/channel/UC3UzBZL-5VRkgoFYinf4kGg> [Actualizado a julio de 2022]

American Cancer Society, “¿Cómo entender el cáncer avanzado, el cáncer metastásico y la metástasis en los huesos?”, *Cáncer avanzado, cáncer metastásico y metástasis en los huesos*, 15 de diciembre de 2016, s/l, consúltase en: <https://www.cancer.org/es/tratamiento/como-comprender-su-diagnostico/cancer-avanzado/que-es.html> [Actualizado a julio de 2022]

Amnistía Internacional y Access Now, The Toronto Declaration: Protecting the rights to equality and non-discrimination in Machine Learning systems: <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/> [Actualizado a julio de 2022]

Angulo Eél, María, “Errores médicos, las sentencias de muerte de 2,6 millones de pacientes por año” en France24, del 15 de septiembre de 2019, consulte en: <https://www.france24.com/es/20190914-errores-medicos-oms-muerte-salud> [Actualizado a julio de 2022]

Agrawal, Dakshi, “AI learns the language of business: Behind the scenes at IBM’s NLP Innovation Event”, *IBM-Watson*, publicado el 17 de marzo de 2020 en el sitio: <https://www.ibm.com/blogs/watson/2020/03/ai-learns-the-language-of-business-behind-the-scenes-at-ibms-nlp-innovation-event/> [Actualizado a julio de 2022]

AI NOW, *Algorithmic Accountability Policy Toolkit*, s/l, 2018. Consúltese en: <https://ainowinstitute.org/aap-toolkit.pdf> [Actualizado a julio de 2022]

Aizenberg, Evgeni y Jeroen van den Hoven. “Designing for Human Rights in AI”, *Big Data & Society*, julio 2020, doi:[10.1177/2053951720949566](https://doi.org/10.1177/2053951720949566).

Ajunwa, Ifeoma, “The ‘Black Box’ at Work”, en *Big Data & Society*, julio 2020, <https://doi.org/10.1177/2053951720938093>. [Actualizado a julio de 2022]

Allen, Kate, “Computer beats Atari in artificial intelligence breakthrough”, *Toronto Star*, febrero 25, 2015. Consúltese en: <https://www.thestar.com/news/world/2015/02/25/computer-beats-atari-in-artificial-intelligence-breakthrough.html> [Actualizado a julio de 2022]

-----, “How a Toronto professor’s research revolutionized artificial intelligence”, *Toronto Star*, abril 17, 2015. Consúltese en <https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html> [Actualizado a julio de 2022]

Anders Chydenius Foundation, *The World’s First Freedom of Information Act*. Anders Chydenius Foundation Publications 2, s/l, 2016. https://www.chydenius.net/tiedostot/worlds_first_foia.pdf [Actualizado a julio de 2022]

AP-Bloomberg, “Las 3 grandes tecnológicas que negarán software de reconocimiento facial a la policía tras caso Floyd”, *El Financiero*, 11 de junio de 2020, en <https://www.elfinanciero.com.mx/tech/las-3-grandes-tecnologicas-que-negaran-software-de-reconocimiento-facial-a-policia-tras-muerte-de-floyd/> [Actualizado a julio de 2022]

Aperioncare, Inventions that drive modern life expectancy, aperioncare, en <https://aperioncare.com/blog/inventions-life-expectancy/> [Actualizado a julio de 2022]

Aris Papakonstantinou, Haridimos Kondylakis, y Emmanouil Marakakis, “INTEGRA: a web-based differential diagnosis system combining multiple knowledge bases”, In Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '20), Association for Computing Machinery, New York, 2020, article 45. DOI: <https://doi.org/10.1145/3389189.3397980> [Actualizado a julio de 2022]

Vara Horna, Arístides Alfredo, *¿Cómo evaluar la rigurosidad científica de las tesis doctorales? Criterios teóricos y metodológicos aplicados a la educación*, Perú, Universidad de San Martín de Porres, 2010.

Aristóteles, *Moral a Nicómaco*, Alianza Editorial, Madrid, 2005.

Argimón Pallàs, J. Ma., Guarga Rojas, A., Jiménez Villa, J., Medicina basada en la evidencia: guías y protocolos, descargado por la Universidad Nacional Autónoma de México, de ClinicalKey.es por Elsevier, junio 25, 2020, consúltese en: <http://dsp.facmed.unam.mx/wp-content/uploads/2016/06/U-2-Anexo-1A-Argim%C3%B3n-J-Medicina-Basada-en-evidencias-gu%C3%ADas-y-protocolos.pdf> [Actualizado a julio de 2022]

Asimov, Isaac, *Yo, robot*, segunda edición, México, Editorial Sudamericana, 2016.

Ashby, W. R., *Design for a brain*, Wiley, New York, 1952.

Association for Computing Machinery, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, 2020. Consúltese en <https://dl.acm.org/doi/proceedings/10.1145/3351095> [Actualizado a julio de 2022]

Association for Computing Machinery, *Statement on Algorithmic Transparency and Accountability Public Policy Council*, ACM Europe Policy Committee, aprobada el 12 de enero de 2017. Consúltese en: http://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf [Actualizado a julio de 2022]

Ayllón, José Ramón, *Antropología filosófica*, Ariel, Barcelona, 2014.

Bajo Fernández, Miguel y Bacigalupo Saggese, Silvina, Curso de derecho penal económico, Editorial Universitaria Ramón Areces, Madrid: 2010.

Bajwa, Junaid “AI & Health: We rise by lifting others”, en Data Science & Law Forum 3.0, 2021. Consúltense en: https://datasciencelawforum.eu/?utm_campaign=833886_Thank%20you%20for%20joining%20the%20Data%20Science%20%26%20Law%20Forum%203.0&utm_medium=dotdigital&utm_source=Forum%20Europe&utm_i=4SEG%2CHVFI%2C4O3GPN%2C246OF%2C1 [Actualizado a julio de 2022]

Barranco Fragoso, Ricardo, ¿Qué es el Big Data? Todos formamos parte de ese gran crecimiento de datos, IBM Corporation, 2012. <https://developer.ibm.com/es/technologies/data-science/articles/que-es-big-data/> [Actualizado a julio de 2022]

Bartneck, Christoph, Wagner Sean, Welsh, Christophütge Alan, *An Introduction to Ethics in Robotics and AI*, Springer, 2021.

Bautista-Hernández, María Yicel y Medina-Villaseñor, Efraín A., “Síndrome de compresión medular por cáncer”, *Gaceta Mexicana de Oncología*, septiembre 2011. <https://www.elsevier.es/es-revista-gaceta-mexicana-oncologia-305-articulo-sindrome-compresion-medular-por-cancer-X1665920111894309> [Actualizado a julio de 2022]

Bentham, Jeremy, Panopticon; or, the inspection-house: containing the idea of a new principle of construction applicable to any sort of establishment, in which persons of any description are to be kept under inspection; and in particular to penitentiary-houses, lazarettos, hospitals and schools: with a plan of management adapted to the principle: in a series of letters, written in the year 1787, from crecheff in White Russia, to a friend in England, Londres, 1791.

-----, *Tratado de las pruebas judiciales*. [Obra extraída de los manuscritos de M. Jeremías BENTHAM], jurisconsulto inglés, escrita en francés, por Estevan DUMONT, Vocal del Consejo representativo y soberano de Ginebra. Trad. C. M. V. Tomo I y II, París, Bossange Frères, 1825. [La obra original es de 1823, París]

Benjamins, S., Dhunnoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digit. Med.* 3, 118 (2020). <https://doi.org/10.1038/s41746-020-00324-0> [Actualizado a julio de 2022]

Berlin, Isahía, *Freedom and Its Betrayal: Six Enemies of Human Liberty - Updated Edition*, Princeton University Press, Oxford, 1952.

Bernal Pinzón, Jesús, “Imputabilidad, culpabilidad y responsabilidad”, *Nuevo Foro Penal*, no. 56, abril 1992, p. 262. <https://publicaciones.eafit.edu.co/index.php/nuevo-foro-penal/article/download/4033/3287> [Actualizado a julio de 2022]

Betti, Emilio, *Teoría generale della Interpretazione*, Milano, Dott. A. Giuffré Editore, 1955.

Bernate Ochoa, Francisco, *Imputación objetiva y responsabilidad penal médica*, Universidad del Rosario, Bogotá, 2010

Beuchot, Mauricio, *Elementos de Semiótica*, Editorial Surge, México, 2001.

Board of Governors of the Federal Reserve System, “The Fed -Supervisory Letter SR 11-7 on Guidance on Model Risk Management”, 2011 <https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm> [Actualizado a julio de 2022]

Boecio, *La consolación de la Filosofía*, Porrúa, México, 2004.

Botelho, Bridget “Cognitive computing”, *TechTarget*, julio 2018, en <https://searchenterpriseai.techtarget.com/definition/cognitive-computing> [Actualizado a julio de 2022]

- Brodwin, Erin, "Google expands controversial pilot project using patient data", *STAT*, febrero 23, 2021, en: <https://www.statnews.com/2021/02/23/google-project-nightingale-care-studio/> [Actualizado a julio de 2022]
- Brown, J. M., *et al.*, "Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks". *JAMA Ophthalmol*, volume 136, 2018. <https://doi.org/10.1001/jamaophthalmol.2018.1934> [Actualizado a julio de 2022]
- Burns, Ed, "IBM Watson supercomputer", en TechTarget, junio de 2018, https://searchenterpriseai.techtarget.com/definition/IBM-Watson-supercomputer?_gl=1*1pwzd5r*_ga*MTM2NTMyNjI2OC4xNjIwODg1OTI5*_ga_RRBYR9CGB9*MTYyMDg4NTkyOC4xLjEuMTYyMDg4NTk0MS4w&_ga=2.267401990.963224639.1620885929-1365326268.1620885929 [Actualizado a julio de 2022]
- Burrell, Jenna, "How the Machine 'Thinks': Understanding Opacity in *Machine Learning Algorithms*" en *Big Data & Society*, junio 2016. <https://doi.org/10.1177/2053951715622512>. [Actualizado a julio de 2022]
- Cabañas García, Juan Carlos, *La valoración de las pruebas y su control en el proceso civil. Estudio dogmático y jurisprudencial*, Trivium, Madrid 1992.
- Caiotlin Cheadle, "Here's How Tech is Disrupting the Traditional Healthcare Market", *Visualcapitalist*, diciembre 13, 2016, en <https://www.visualcapitalist.com/tech-disrupting-healthcare-market/> y How Cloud Technology Is Changing Healthcare en: <https://www.mbahealthcaremanagement.org/topics/how-technology-is-changing-healthcare/> [Actualizado a julio de 2022]
- Calvo, Manuel, *Filosofía para la era digital*, Almuzara, Madrid, 2018.
- Caplan Robyn, Donovan, Joan, Hanson, Lauren and Matthews, Jeanna, *Algorithmic Accountability: A Primer*, Data & Society, Washington, 2018. Consúltase en: https://datasociety.net/wp-content/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL-4.pdf [Actualizado a julio de 2022]

Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, *et al.*, “Human-centered tools for coping with imperfect algorithms during medical decision-making”, en *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 4, 2019.

Central Digital and Data Office, Guidance for public sector organizations on how to use data appropriately and responsibly when planning, implementing, and evaluating a new policy or service, 16 de septiembre de 2020. Consúltense en: <https://www.gov.uk/government/publications/data-ethics-framework> [Actualizado a julio de 2022]

Cerdio, Jorge, “Human Rights in the Era of Artificial Intelligence”, Fletcher School, 2021. Conferencia impartida el 9 de marzo de 2021 [Actualizado a julio de 2022] <https://www.youtube.com/watch?v=zK1LW9CcdxM> [Actualizado a julio de 2022]

-----, “Human Rights, Democracy and Rule of Law Impact Assessment of AI systems”, el Proyecto de la Unión Europea”, Primer Congreso de Lógica, Epistemología y Ética de la Inteligencia Artificial organizado por la Facultad de Filosofía y la Facultad de Ingeniería de la Universidad Panamericana, México, 2021. Consúltense: <https://fb.watch/v/2VFm7bJet/> [Actualizado a julio de 2022]

-----, *Asumir y no posponer. La educación efectiva, y también digital*, s/l., s/e., 2019.

Cervantes, JA., López, S., Rodríguez, LF. et al. Artificial Moral Agents: A Survey of the Current Status. *Sci Eng Ethics* 26, 2020. <https://doi.org/10.1007/s11948-019-00151-x> [Actualizado a julio de 2022]

Chen, Angela, “IBM’s Watson Gave Unsafe Recommendations For Treating Cancer”, *Verge*, 26 de julio de 2018.

Coalición IA2030 Mx, *Agenda Nacional Mexicana de Inteligencia Artificial*, México, CMinds. <https://www.ia2030.mx/> [Actualizado a julio de 2022]

Coleman, Jules L., "The Morality of Strict Tort Liability", 18 Wm. & Mary L. Rev. 259, 1976, Consúltese en: <https://scholarship.law.wm.edu/wmlr/vol18/iss2/3> [Actualizado a julio de 2022]

Comisión de Bruselas, Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial Creado por la Comisión Europea, *Directrices Éticas para una IA fiable*, Bruselas, 2019.

Comisión Europea, *Libro Blanco sobre la inteligencia artificial -un enfoque europeo orientado a la excelencia y la confianza*, Bruselas, 2020. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf [Actualizado a julio de 2021]

Comité de Derechos Económicos, Sociales y Culturales, *Observación general N.º 14 sobre el derecho a la salud*, aprobada por el Comité de Derechos Económicos, Sociales y Culturales, 2020. Consúltese en: <https://www.acnur.org/fileadmin/Documentos/BDL/2001/1451.pdf> [Actualizado a julio de 2022]

Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights. Recommendation*, Council of Europe, Bruselas, 2019. Consúltese en: <https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights> [Actualizado a julio de 2021]

-----, *Algorithmic Accountability Policy Toolkit*, Octubre 2018.

Consejo Nacional de Salud Pública, A. C., *Lex Artis ad Hoc. Especialidades de Salud Pública, Epidemiología y Medicina Preventiva*, México, 2018. Consúltese en: [lexartis.cdr \(cnsaludpublica.com.mx\)](http://lexartis.cdr.cnsaludpublica.com.mx) [Actualizado a julio 2021]

Cooper, Laura y Lombardo, Cara, IBM Explores Sale of IBM Watson Health, The Wall Street Journal, febrero 18, 2021, en <https://www.wsj.com/articles/ibm-explores-sale-of-ibm-watson-health-11613696770> [Actualizado a julio de 2022]

Copeland, Jack, *Inteligencia artificial*, Alianza Universidad, Madrid, 1996.

Copi Irving M. y Cohen, Carl, *Introducción a la lógica*, Limusa, México, 2013.

Cuairán, Javier, “La responsabilidad penal de los nuevos entes de inteligencia artificial”, El país, sección Cinco días, 1 de junio de 2021, versión digital. Consúltase en: https://cincodias.elpais.com/cincodias/2021/05/31/legal/1622463707_166390.html

Cuartas Báez, Rafael Y., “La Blockchain como elemento de gobernanza, confianza, Ética y autoregulación de la IA” en XVI Conferencia del Ciclo Sistemas Inteligentes y Ética, Academia nacional de Ciencias de Buenos Aires, 11 de noviembre 2020. Cfr. [15º Conferencia del Ciclo Sistemas Inteligentes y Ética - CETI - ANCBA](#) [Actualizado a julio de 2022]

Dastin, Jeffrey, “Amazon scraps secret AI recruiting tool that showed bias against women”, *Reuters*, 10 de octubre de 2018, en <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> [Actualizado a julio de 2022]

Data Science & Law Forum
https://datasciencelawforum.eu/?utm_campaign=833886_Thank%20you%20for%20joining%20the%20Data%20Science%20%26%20Law%20Forum%203.0&utm_medium=dotdigital&utm_source=Forum%20Europe&dm_i=4SEG%2CHVFI%2C4O3GPN%2C246OF%2C1 [Actualizado a julio de 2022]

Davies, Jessica, “After GDPR, The New York Times cut off ad exchanges in Europe — and kept growing ad revenue”. *DIGIDAY*, enero 16, 2019. <https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/> [Actualizado a julio de 2022]

Davies, Jessica, “Axel Springer counters Google with its own consent management tool”, *DIGIDAY*, mayo 30, 2018. <https://digiday.com/media/axel-springer-counters-google-consent-management-tool/> [Actualizado a julio de 2022]

De George, R. T., *The ethics of information technology and business*, Malden, Blackwell Publishing, 2003.

Demarta, Jean-Christophe, The Editorial Board. America's Privacy Settings Are Wrong. Section SR, *The New York Times*. New York. Marzo 7, 2021

Desjardins, Jeff, *The 50 Most Important Life-Saving Breakthroughs in History*, marzo 26, 2018, consúltese en: <https://www.visualcapitalist.com/50-important-life-saving-breakthroughs-history/> [Actualizado a julio de 2022]

-----, *The 6 Forces Transforming the Future of Healthcare*, *Visual capitalist*, julio 12, 2018, consúltese en: <https://www.visualcapitalist.com/6-forces-transforming-future-healthcare/> [Actualizado a julio de 2022]

Destin, J., "Amazon scraps secret AI recruiting tool that showed bias against women", *Reuters*, octubre 9, 2018.

Dharmendra S. Modha, Rajagopal Ananthanarayanan, Steven K. Esser, Anthony Ndirango, Anthony J. Sherbondy, and Raghavendra Singh. 2011. "Cognitive computing". *Commun. ACM* 54, 8, agosto de 2011. DOI: <https://doi.org/10.1145/1978542.1978559> [Actualizado a julio de 2022]

Diakopoulos, Nicholas, *et. al.*, *Principles for accountable algorithms and a social impact statement for algorithms*, 2020.

Diarioti, "IBM compra dos startups y crea *Watson Health*, unidad de negocio orientada al sector de la salud", *Diarioti*, abril 14, 2015. Consúltese: <https://diarioti.com/ibm-compra-dos-startups-y-crea-watson-health-unidad-de-negocio-orientada-al-sector-de-la-salud/87178> [Actualizado a julio de 2022]

-----, Suspenden a IBM Watson por recomendar "tratamiento mortal" para pacientes de cáncer, *Diarioti*, noviembre 1, 2017, consúltese en: <https://diarioti.com/suspenden-a-ibm-watson-por-recomendar-tratamiento-mortal-para-pacientes-de-cancer/106013> [Actualizado a julio de 2022]

Diéguez, Antonio, *Transhumanismo. La búsqueda tecnológica del mejoramiento humano*, Herder, Barcelona, 2017.

Dieterich Steffan, Heinz, *Nueva Guía para la investigación científica*, Universidad de Ciencias y Humanidades, Lima, 2008.

Dietterich, Thomas G, “*Machine Learning*”, *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., Londres, 2013. <https://dl.acm.org/doi/pdf/10.5555/1074100.1074563> [Actualizado a julio de 2022]

Dignum, Virginia, *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*, Springer, 2019, Suiza.

Dimitrios Zikos. “A Framework to Design Successful Clinical Decision Support Systems”, en Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '17), Association for Computing Machinery, New York, 2017. DOI:<https://doi.org/10.1145/3056540.3064960> [Actualizado a julio de 2022]

Dirección General de Derechos Humanos y Estado de Derecho del Consejo de Europa, Facial recognition: strict regulation is needed to prevent human rights violations, News 2020-2021, 28 de enero de 2021. Consúltese en: <https://www.coe.int/en/web/human-rights-rule-of-law/-/facial-recognition-strict-regulation-is-needed-to-prevent-human-rights-violations> [Actualizado a julio de 2022]

Döhring, Erich, *La prueba, su práctica y apreciación*, trad. Banzhaf, Buenos Aires 1964.

Dokyun Lee, Joy Lu, Kim, Tae Wan y David Danks, “Good Explanation for Algorithmic Transparency. Proceedings of the AAAI/ACM” *Conference on AI, Ethics, and Society*, Association for Computing Machinery, New York, número 93, 2020. en DOI: <https://doi.org/10.1145/3375627.3375821> [Actualizado a julio de 2022]

- Doshi-Velez, Finale y Been Kim. “Towards A Rigorous Science of Interpretable *Machine Learning*.” *arXiv: Machine Learning*, 2017. <https://arxiv.org/pdf/1702.08608.pdf> [Actualizado a julio de 2022]
- Doucette, Marc, “Why Big Tech is Plotting an Invasion of the Healthcare Market”, *Visual capitalist*, abril 24, 2018. Consúltese en: <https://www.visualcapitalist.com/big-tech-invading-healthcare-market/> [Actualizado a julio de 2022]
- Douglas L. McGee, *Errores cognitivos en la toma de decisiones médicas*, DO, Emergency Medicine Residency Program, Albert Einstein Medical Center, 2018. Consúltese en: <https://www.msmanuals.com/es/professional/temas-especiales/toma-de-decisiones-m%C3%A9dicas/errores-cognitivos-en-la-toma-de-decisiones-m%C3%A9dicas> [Actualizado a julio de 2022]
- Drozdal, Jaimie, Weisz, Justin, et. al., “Trust in AutoML: exploring information needs for establishing trust in automated *Machine Learning* systems” en *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020.
- Dwork, Cynthia, Hardt, Moritz, et. Al. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12)*. Association for Computing Machinery, New York, NY, USA, 214–226. DOI: <https://doi.org/10.1145/2090236.2090255> [Actualizado a julio de 2022]
- Dworkin, Ronald, *Justicia para erizos*, FCE, México, 2014.
- Eco, Umberto, *Cómo se hace una tesis*, Gedisa, Barcelona, 2010.
- Eiben A.E. y Smith J.E., *Introduction to Evolutionary Computing*, Springer, Berlin, 2013.
- Endsley, M. R., Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37, 1997, 32–64. <https://doi.org/10.1518/001872095779049543>
- Engisch, Karl, *La teoría de la libertad de la voluntad en la actual doctrina filosófica del derecho penal*, Editorial B de F, Buenos Aires, 2006.

Evis Echandia, Hernando, *Teoría General de la Prueba Judicial*, Tomos 1 a 6, Víctor P. de Zavalía Editor, Buenos Aires. (diversas fechas).

Fails, Jerry Alan y Olsen Jr., Dan R, “Interactive Machine Learning” en *Proceedings of the 8th international conference on Intelligent user interfaces*. ACM, 2003.

Feigenbaum, Edward A. y Klahr, Philip, “Expert Systems”, *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., Londres, 2013, pp. 684–689. Consúltese en: <https://dl.acm.org/doi/10.5555/1074100.1074389> [Actualizado a julio de 2022]

Ferrajoli, Luigi, *Derecho y razón. Teoría del Garantismo penal*, Trotta, Madrid, 2009.

-----, *Principia iuris. Teoría del derecho y de la democracia*. Tomo I. Teoría del derecho, Trotta, Madrid, 2011.

Fernández Ardavín, Bernardo, *Ética de la Dirección. Decisión con libertad*, IPADE-Ediciones Ruz, México, 2006.

Fernández Cantón, Sonia B., *Boletín CONAMED- OPS*, México, marzo – abril 2016.
Consúltese en:
http://www.conamed.gob.mx/gobmx/boletin/pdf/boletin5/acto_medico.pdf
[Actualizado a julio de 2022]

Ferrer Beltrán, Jordi, *La valoración racional de la prueba*, Marcial Pons, Madrid, 2007.

Fix-Zamudio, Héctor, *Metodología, Docencia e Investigación Jurídicas*, Porrúa, México, 2007.

Foucault, Michel, “Incorporación del hospital en la tecnología moderna” en *Educ Med Salud*, Vol. 12, no. 1, 1978.

Friedberg, Mark W, Chen, Peggy G., et. al., “Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy”, *Rand health quarterly*, número 3, 4, 2014.

Friedman, Thomas L., *Thank You for Being Late*. Farrar, Straus and Giroux. Edition Kindle, New York, 2016.

Frisch, Wolfgang, *Comportamiento típico e imputación del resultado*, Marcial Pons, Madrid, 2004.

Frosini, Vittorio, *Cibernética, Derecho, Internet y Sociedad*, Lex Ediciones Olejnik, Santiago, 2019.

-----, Informatica e diritto, *XXXIV Annata*, vol. XVII, 2008, n. 1-2, en: frontespizio e dedica.qxd (cnr.it)

Fuster, Joaquín "M., *Cerebro y Libertad. Los cimientos cerebrales de nuestra capacidad para elegir*, Ediciones Culturales Paidós, México, 2019.

Gadamer, Hans-Georg, *Verdad y método*, Octava Edición, Tomos I y II, Ediciones Sígueme, Salamanca 1999.

García Serrano, Alberto, *Inteligencia artificial: Fundamentos, práctica y aplicaciones*, Alfaomega, México, 2016.

Gartner, "Magic Quadrant for Data Science and *Machine Learning* Platforms", Gartner, publicado el 1 de marzo 2021. Consúltase en: <https://www.gartner.com/doc/reprints?id=1-25DOZD29&ct=210304&st=sb>
[Actualizado a julio de 2022]

Gascón Abellán, Marina, *Los hechos en el derecho. Bases argumentales de la prueba*, Marcial Pons, Madrid-Barcelona, 2004.

Gigerenzer, Gerd, *Decisiones instintivas*, Ariel, Edición de Kindle, 2018.

Gilpin, L.H., Testart, C., Fruchter, N., y Adebayo, J., *Explanations to Society*, ArXiv, abs/1901.06560. [1901.06560v1.pdf \(arxiv.org\)](https://arxiv.org/abs/1901.06560v1) 2019. [Actualizado a julio de 2022]

Gimbernát Ordeig, Enrique, “Casualidad, omisión e imprudencia” en *Anuario de derecho penal y ciencias penales*, tomo 47, Fasc/Mes 3, Madrid, 1994.

-----, “En defensa de teoría de la imputación objetiva contra sus detractores y, también, contra algunos de sus partidarios”, *Anuario de Derecho Penal y Ciencias Penales*, Tomo 73, Fasc. t. 1, 2020, Madrid.

Glancy, Dorothy J., “The Invention of the Right to Privacy”, *Arizona Law Review*, volumen 21, número 1, 1979.
<https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?referer=https://en.wikipedia.org/&httpsredir=1&article=1318&context=facpubs> [Actualizado a julio de 2022]

Gómez Mont, Constanza, May Del Pozo, Claudia, Martínez Pinto, Cristina; Martín del Campo Alcocer, Ana Victoria. *La Inteligencia Artificial al Servicio del Bien Social en América Latina y el Caribe: panorámica regional e instantáneas de doce países*, Banco Interamericano de Desarrollo, México, 2020. https://7be0258a-f55e-4968-87a4-4c067a3d1160.usrfiles.com/ugd/7be025_bbfdb0a0224c43a2ad371b87b66c2eae.pdf [Actualizado a julio de 2022]

Gong, Bo, Nugent, James P., *et. al.*, “Influence of Artificial Intelligence on Canadian Medical Students Preference for Radiology Specialty: A National Survey Study” en *Academic radiology*, número 26, 4, 2019.

González Alcántara, Juan Luis, *La responsabilidad civil de los médicos*, Porrúa, México, 2009.

González Amuchástegui, Jesús, *Autonomía, Dignidad y Ciudadanía. Una teoría de los Derechos Humanos*, Tirant lo Blanch, Valencia, 2004.

Ghosh, Imán, Visualizing the Healthtech Revolution, en Visual Capitalist, marzo 26, 2019, consúltese: <https://www.visualcapitalist.com/visualizing-healthtech-revolution/> [Actualizado a julio de 2022]

Granero, Horacio, *El orden público tecnológico*, Educa, Buenos Aires, 2003.

-----, ¿Que límites éticos se plantean hoy un hacedor de programas inteligentes? en el Ciclo de Conferencias sobre Sistemas Inteligentes y Ética, Academia Nacional de la Ciudad de Buenos Aires, 12 de agosto de 2020 [2º conferencia del Ciclo "Sistemas Inteligentes y Ética" - CETI - ANCBA - 12/08/20](#) [Actualizado a julio de 2022]

Gupta, Kousini, “Actus Non Facit Reum Nisi Hombro Sit Rea” - *Legal Maxim*, en *Law Times Journal*, 26 de junio de 2019. Consúltese en <https://lawtimesjournal.in/actus-non-facit-reum-nisi-mens-sit-rea/> [Actualizado a julio de 2022]

Gutiérrez y González, Ernesto, *Derecho de las obligaciones*, XXIII Edición, Porrúa, 2020.

Guyatt G, Cairns J, Churchill D, *et al.*, “Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine”, *JAMA*, volume 268 (17), 1992. doi:10.1001/jama.1992.03490170092032 [Actualizado a julio de 2022]

Haider Khalaf Jabbar and Rafiqul Zaman Khan, “Survey on Development of Expert System from 2010 to 2015”, en *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (ICTCS '16)*. Association for Computing Machinery, New York, article 130, 1–7. DOI: <https://doi.org/10.1145/2905055.2905190> [Actualizado a julio de 2022]

Hans Jonas, *El principio de responsabilidad. Ensayo de una Ética para la civilización tecnológica*, Herder, Barcelona, 1995.

Hans Welzel, *El nuevo sistema del derecho penal. Una introducción a la teoría de la acción finalista*, Editorial B de F, Buenos Aires, 2004; y también *Derecho Penal Alemán. Parte General*, Editorial Jurídica de Chile, Santiago de Chile, 1970.

Hart, Herbert Lionel Adolphus, *Punishment and Responsibility*, Clarendon Press, Oxford, 1968. [En 2019, la editorial Marcial Pons publica la edición traducida por Jacobo

Barja de Quiroga y León García-Comendador, Alonso, *Castigo y responsabilidad* (Madrid)]

Hayek, F. A., *The sensory order*, Chicago Press, Chicago, 1952.

Hebb, D. O., *The organization of behavior*. Wiley, New York, 1949.

Henk Ten Have. 2000. Re-evaluating professional autonomy in health care. *Theoretical medicine and bioethics* 21, 5 (2000), 503–513.

Hernández Sánchez, José Luis, *Curso Introducción a la Inteligencia Artificial* (Impartido por: Rav Ahuja, AI and Data Science Program Director IBM) s/e, México, 2020.

-----, José Luis, *Sistemas Expertos Decisionales*, Universidad Anáhuac Online, Diplomado Ciberjusticia, México, 2020.

Hernández-Sampieri Roberto y Mendoza Torres, Christian Paulina, *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*, Mc Graw Hill, México, 2019.

Hernández-Sampieri, Roberto, Méndez Valencia, Sergio, *et. al.*, *Fundamentos de investigación*, Mc Graw Hill, México, 2017.

Herranz, Arantxa, “20 años después quiero volver a programar: qué grandes cambios ha habido y por dónde empezar”, XATAKA, 22 de junio de 2020, en <https://www.xataka.com/otros/20-anos-despues-quiero-volver-a-programar-que-grandes-cambios-ha-habido-donde-empezar> [Actualizado a julio de 2022]

Herranz, Julián, *Dios y Audacia. Mi juventud junto a Josemaría*, segunda edición, Ediciones RIALP, S.A., Madrid, 2011.

Hidalgo Pérez, Montse, La Unesco pone los cimientos de la ‘Declaración Universal de la inteligencia artificial’ | Tecnología | EL PAÍS (elpais.com), 17 de septiembre de 2020, Disponible en: <https://elpais.com/tecnologia/2020-09-17/la-unesco-pone-los->

[cimientos-de-la-declaracion-universal-de-la-inteligencia-artificial.html](#) [Actualizado a julio de 2022]

-----, “En cuestiones tecnológicas, la Ética lo es todo y debe prevalecer en todo” | Tendencias | EL PAÍS Retina (ampproject.org), 27 de diciembre de 2020, Disponible en: https://elpais.com/retina/2020/12/22/tendencias/1608655621_852613.html [Actualizado a julio de 2022]

Hill, Caroline, “Artificial Intelligence in Law – The State of Play in 2015?”, *Legal Technology*, 3 de noviembre de 2015. Consúltase en: <https://legaltechnology.com/artificial-intelligence-in-law-the-state-of-play-in-2015/> [Actualizado a julio de 2022]

Hinton, G. E., *Relaxation and its role in vision*. Tesis de doctorado, Universidad de Edimburgo, Edimburgo, 1978. Consúltase el texto original en: <https://era.ed.ac.uk/handle/1842/8121> [Actualizado a julio de 2022]

Hipócrates, Tratados hipocráticos, Tomo I y VIII, Gredos, Madrid, 1983.

Holmes, Oliver Wendell, *The Common Law*, 1881, en https://www.gutenberg.org/files/2449/2449-h/2449-h.htm#link2H_4_0002 [Actualizado a julio de 2022]

Hornig S., et al Creating an automated trigger for sepsis clinical decision support at emergency department triage using Machine Learning. PLoS ONE (2017). Sheth, D. & Giger, M. L. Artificial intelligence in the interpretation of breast cancer on MRI. J. Magn. Reson. Imaging 51, 1310–1324, <https://doi.org/10.1002/jmri.26878> (2019). [Actualizado a julio de 2022]

Hu, Margaret. “Cambridge Analytica’s Black Box” en *Big Data & Society*, julio 2020. <https://doi.org/10.1177/2053951720938091>. [Actualizado a julio de 2022]

IBM News Room, “BNP Paribas Signs an Agreement with IBM Services to further deploy its Cloud Strategy”, 22 de enero de 2019. Consúltase en:

<https://newsroom.ibm.com/2019-01-22-BNP-Paribas-Signs-an-Agreement-with-IBM-Services-to-further-deploy-its-Cloud-Strategy> [Actualizado a julio de 2022]

IBM, “A billion people a day. Millions of elevators. No room for downtime. KONE and Watson IoT analyze data in elevators and escalators around the world to keep people moving smoothly, safely, and efficiently”, *Watson stories*, s/f. Consúltese en: <https://www.ibm.com/watson/stories/ca-en/kone-with-watson.html> [Actualizado a julio de 2022]

IBM, “Abu Dhabi National Oil Company (ADNOC)”, Case studies, s/f. Consúltese en: <https://www.ibm.com/case-studies/abu-dhabi-national-oil-company-adnoc> [Actualizado a julio de 2022]

IBM, “Customer response in seconds, not minutes” *Watson stories*, s/f. Consúltese en: https://www.ibm.com/watson/stories/bradesco?mhsrc=ibmsearch_a&mhq=Caso%201%20Bradesco [Actualizado a julio de 2022]

IBM, “How It Works: IBM Watson”, IBM, 21 de noviembre de 2014. Consúltese en: <https://www.youtube.com/watch?v=AtdJ1DGJjXA&t=138s> [Actualizado a julio de 2022]

IBM, “IBM Watson services for Core ML”, *Watson stories*, s/f. Consúltese en: [IBM Watson Services for Core ML | IBM](#) [Actualizado a julio de 2022]

IBM, “Pure Application System”, PureSystems. EExpertise. Pure and Simple, s/f. Consúltese en: https://www.ibm.com/ibm/puresystems/uy/es/pf_pureapplication.html [Actualizado a julio de 2022]

IBM, “Supporting cancer research and treatment”, *Watson Health*, s/f. Consúltese en: <https://www.ibm.com/watson-health/solutions/cancer-research-treatment> [Actualizado a julio de 2022]

IBM, “What happens when AI stops playing games?”, *The Verge*, 6 de mayo de 2020, consúltese en: <https://www.youtube.com/watch?v=7pHaNMdWGsk> [Actualizado a julio de 2022]

IBM, Archives, “Valuable resources on IBM's history”, *IBM Archives*, s/f. Consúltese en el sitio: <https://www.ibm.com/ibm/history/index.html> [Actualizado a julio de 2022]

IEEE Spectrum, “IBM Watson, health thyself: How IBM overpromised and underdelivered on AI health care”, vol. 56, no. 4, abril 2019. doi: 10.1109/MSPEC.2019.8678513. <https://ieeexplore.ieee.org/document/8678513> [Actualizado a julio de 2022]

INCIBE, *Acuerdo pactado, contrato firmado. ¡Protege tu empresa!*, Instituto Nacional de Ciberseguridad, 15 de diciembre de 2020, Leon, en <https://www.incibe.es/protege-tu-empresa/blog/acuerdo-pactado-contrato-firmado-protege-tu-empresa> [Actualizado a julio de 2022]

IONOS, Reglamento ePrivacy: ¿qué hay que saber?. Digital Guide, Marzo 3, 2021. <https://www.ionos.mx/digitalguide/paginas-web/derecho-digital/eprivacy-reglamento-sobre-privacidad-electronica-en-la-eu/> [Actualizado a julio de 2022]

iProUP, “Microsoft no venderá su software de reconocimiento facial a la policía: por qué tomó esta decisión”, *iProUP*, 12 de junio de 2020, en: <https://www.iproup.com/innovacion/14477-por-que-microsoft-no-vendera-su-software-de-reconocimiento-facial-a-la-policia> [Actualizado a julio de 2022]

Ipsos MORI, *Public views of Machine Learning: findings from public research and engagement*, Londres, Ipsos Mori, 2017. <https://royalsociety.org/-/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf> [Actualizado a julio de 2022]

Ishiguro, Kazuo, *Klara y el Sol*, Editorial Anagrama, México, 2021.

Jakobs, Günther, *La imputación objetiva en el derecho penal*, Editorial Ad-Hoc, Buenos Aires, 1997.

Johnson, D.G., Technology with No Human Responsibility?, J Bus Ethics, 127, 2015.
<https://doi.org/10.1007/s10551-014-2180-1> [Actualizado a julio de 2022]

Joyanes Aguilar, Luis, *Fundamentos de programación. Algoritmos, estructura de datos y objetos*, McGraw Hill, Cuarta Edición, Madrid, 2004.

-----, *Fundamentos de programación. Algoritmos, estructura de datos y objetos*, McGraw Hill, Quinta Edición, México, 2020.

Joyanes, Luis, *Big Data - Análisis de grandes volúmenes de datos en organizaciones*, Alfaomega. Edición de Kindle.

Kant, Emmanuel, *Fundamentación de la Metafísica de las Costumbres*, Trad. Manuel García Morente, Edición de Pedro M. Rosario Barbosa, Puerto Rico, 2007 en:
https://pmrb.net/books/kantfund/fund_metaf_costumbres_vD.pdf [Actualizado a julio de 2022]

Kassam, Ashifa, "Spain's everyday internet warrior who cut free from Google's tentacles",
The Guardian. 13 de mayo, 2014.
<https://www.theguardian.com/technology/2014/may/13/spain-everyman-google-mario-costeja-gonzalez> [Actualizado a julio de 2022]

Kelleher, John D., *Deep Learning, The MIT Press Essential Knowledge series*, Cambridge, The MIT Press, Massachusetts, 2019.

Kizilcec, René F., "How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. Proceedings of the 2016 CHI", *Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, 2016. DOI:
<https://doi.org/10.1145/2858036.2858402> [Actualizado a julio de 2022]

Knoke, William, *Bold New World: The Essential Road Map to the Twenty-First Century*, Kodansha Amer Inc., 1997.

Kostopoulos, Lydia, *Decoupling Human Characteristics from algorithmic capabilities*, The IEEE Standards Association, January 2021.

Krishna, Arvind, *Letter to Congress on Racial Justice Reform*, IBM, 8 de junio de 2020.
Consúltese en: <https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/> [Actualizado a julio de 2022]

Kubat, Misoslav, *An Introduction to Machine Learning*, Springer, Florida, 2017.

Laat, P.B., “Algorithmic Decision-Making Based on *Machine Learning* from Big Data: Can Transparency Restore Accountability?”, *Philos. Technol*, número 31, 2018.
<https://doi.org/10.1007/s13347-017-0293-z> [Actualizado a julio de 2022]

Lacey, Nicola, “Responsibility Without Consciousness” en *Oxford Journal of Legal Studies*, volume 36, número 2, 2016. <https://doi.org/10.1093/ojls/gqv032>
<https://academic.oup.com/ojls/article/36/2/219/2472470>

Lacey, Nicola, *A life of H. L. A. Hart. The nightmare and the noble dream*, Oxford University Press, Oxford, segunda reimpresión, 2012.

Lai, Matthew. “Giraffe: Using Deep Reinforcement Learning to Play Chess”, ArXiv abs/1509.01549, 2015, s/p.

Larrañaga Monjaraz, Pablo, *El concepto de responsabilidad en la teoría del derecho contemporánea*, Tesis para obtener el grado de doctor en Derecho, Director: Manuel Atienza, Universidad de Alicante, Alicante, 1996.

Laudan, Larry, *Verdad, error y proceso penal*, Marcial Pons, Madrid 2013.

LeCun, Y., Bengio, Y., y Hinton, G., “*Deep Learning*”, *Nature*, número 521, 2015.
Consúltese en: <https://doi.org/10.1038/nature14539> [Actualizado a julio de 2022]

Leslie, David, *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*, Zenodo, 2019. <https://doi.org/10.5281/zenodo.3240529> [Actualizado a julio de 2022]

Le Pan, Nicolás, History of pandemics, Visual Capitalist, <https://www.visualcapitalist.com/history-of-pandemics-deadliest/> [Actualizado a julio de 2022]

Levy-Bruhl, Lucien, *L'Idée de Responsabilité*, Paris, Librairie Hachette et Cie, 1884.

Liao, S., "Expert system methodologies and applications - a decade review from 1995 to 2004." en *Expert Syst. Appl.* No. 28, 2005. Consúltese en: http://www.sci.brooklyn.cuny.edu/~kopec/cis718/fall_2005/sdarticle5.pdf [Actualizado a agosto de 2021]

Libet, Benjamin, *Mind Time: The Temporal Factor in Consciousness*, Harvard University Press, 2005.

Llano Cifuentes, Carlos, *Las formas actuales de la libertad*, Trillas, México, 2014, p. 108.

-----, *Ser del hombre y hacer de la organización*, IPADE-Ediciones ECA, México, 2018.

-----, *Dilemas éticos de la empresa contemporánea*, IPADE-Ediciones ECA, México, 2019.

Lombrozo, Tania, "The structure and function of explanations". *Trends in cognitive sciences*, volume 10, 2006. <https://doi.org/10.1016/j.tics.2006.08.004> [Actualizado a julio de 2022]

Lohr, Steve, IBM Buys Truven for \$2.6 Billion, Adding to Trove of Patient Data, The New York Times, febrero 18, 2016, en <https://www.nytimes.com/2016/02/19/technology/ibm-buys-truven-adding-to-growing-trove-of-patient-data-at-watson-health.html> [Actualizado a julio de 2022]

Madeleine, Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction", *Engaging Science, Technology, and Society*, volumen 5, 2019, pp. 40-60. Consúltese en: DOI:10.17351/ests2019.260 [Actualizado a julio de 2022]

Makena Kelly, "The FTC is investigating data collection at YouTube, Facebook, and seven other companies", *The Verge*, 14 diciembre 2020, <https://www.theverge.com/2020/12/14/22174755/ftc-data-privacy-youtube-facebook-reddit-twitter-snapchat-discord-inquiry> [Actualizado a julio de 2022]

Mantelero, Alessandro y Esposito, M.S., "An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems" en *Computer Law & Security Review*, 2021.

Mark, Jade S. Corpuz, Junnel Avestro, Fredilyn B. Calanda, Roselia C. Morco, Jean M. Ángeles, and Carl Matthew S. Alcala, "Rheumatologist: Mobile-based Expert System for Rheumatology Utilizing Fuzzy Logic Algorithm", en *Proceedings of the 2017 International Conference on Computer Science and Artificial Intelligence (CSAI 2017)*, Association for Computing Machinery, New York, 2017. DOI: <https://doi.org/10.1145/3168390.3168428> [Actualizado a julio de 2022]

Martí Andrés, Gabriel, "Sustancia individual de naturaleza racional: el principio personificador y la índole del alma separada" en *Metafísica y Persona. Filosofía, conocimiento y vida*, Año 1 Enero-Junio 2009 Número 1, Universidad de Málaga, pp. 113-119 en: <https://revistas.uma.es/index.php/myp/article/view/2849> [Actualizado a julio de 2022]

Medika Life, "Fifty of the Most Impactful Tech Innovators in Digital Healthcare in 2021", The Medika Life Top 50 Watchlist for digital health companies impacting the healthcare industry in 2021 Fifty of the Most Impactful Tech Innovators in Digital Healthcare in 2021 | Medika Life

Mazeaud Henri, et. al. *Tratado Teórico y Práctico de la Responsabilidad Civil Delictual y Contractual. Compendio*. Tomo I, Buenos Aires, EJEa, 1977.

Mearian, Lucas, *Can anyone afford an IBM Watson supercomputer?*, 2011, <https://www.computerworld.com/article/2513312/can-anyone-afford-an-ibm-watson-supercomputer---yes-.html> [Actualizado a julio de 2022]

Mele, Alfred R., *Libres, por qué a ciencia no ha rebatido la existencia del libre albedrío*, Avarigani, Valencia, 2017.

Méndez Ramírez, Ignacio, et. al., *El protocolo de investigación. Lineamientos para su elaboración y análisis*, Trillas, México, 2004.

Merino, Marcos, “El Premio Turing, para tres de los principales responsables del actual auge de la inteligencia artificial”, *Xataka*, 29 de marzo de 2019. Consúltese en: <https://www.xataka.com/inteligencia-artificial/premio-turing-para-tres-principales-responsables-actual-auge-inteligencia-artificial#:~:text=Geoffrey%20Hinton,-Geoffrey%20Hinton%2C%20brit%C3%A1nico&text=En%201986%20invent%C3%B3%20junto%20a,supervisado%20de%20transformaciones%20de%20imagen%22> . [Actualizado a julio de 2022]

Michalski, Ryszard y Carbonell, Jaime y Mitchell, T., *Machine Learning. an Artificial Intelligence approach*, Tioga Publishing Co., Volume 2, Palo Alto, 1983. https://www.researchgate.net/publication/238582946_Machine_learning_an_Artificial_Intelligence_approach_Volume_2 [Actualizado a julio de 2022]

Microsoft Azure, *How does Microsoft Azure work?*, Microsoft Azure, publicado el 18 de junio de 2018. Consúltese en: <https://www.youtube.com/watch?v=KXkBZCe699A> [Actualizado a julio de 2022]

Minsky, Marvin, *Semantic Information Processing*, MIT Press, Massachusetts 1968.

Mitchell, Tom, *Machine Learning*, Mc Graw Hill Science, s/l, 1997.

-----, *The Discipline of Machine Learning*, School of Computer Science Carnegie Mellon University, Pittsburgh, 2016.

<http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf> [Actualizado a julio de 2022]

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M.A, *Playing Atari with Deep Reinforcement Learning*, 2013. *ArXiv*, *abs/1312.5602*.

Monasterio Astobiza, Aníbal, “Ética algorítmica: Implicaciones Éticas de una sociedad cada vez más gobernada por algoritmos”, *Dilemata*, número 24, mayo 2017.

Monasterio Astobiza, Aníbal, “Ética para máquinas: Similitudes y diferencias entre la moral artificial y la moral humana”, *ILEMATA, Revista Internacional de Éticas Aplicadas*, no. 30, 2019.

Moor, J. H., “The nature, importance, and difficulty of machine ethics”, *IEEE Intelligent Systems*, 21, 4, 2006.

Moore, Andrew W., Shane Shaneman, Martial Hebert, “The AI stack: a blueprint for developing and deploying artificial intelligence”, *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent*, Proc. SPIE 10635, ISR IX, 106350C (4 de mayo de 2018); <https://doi.org/10.1117/12.2309483> [Actualizado a julio de 2022]

Moya, Carlos J, “Prologo”, en Alfred R. Mele, *Libres, por qué la ciencia no ha rebatido la existencia del libre albedrío*, Avarigani, Valencia, 2017.

Moya, Carlos J, *Libre albedrío. Un estudio filosófico*, Cátedra, Madrid, 2017.

Müller, Martin U., “Medical Applications Expose Current Limits of AI”, *Spiegel*, 3 August 2018.

Muñoz Sabaté, Lluís, *Tratado de probática judicial*, Bosch Editor, Barcelona, 1992.

Murray, Silverman M, y Bryan, TJ, *The Quotable Osler*, American College of Physicians, Filadelfia, 2003.

Myers, Andrew, “Stanford’s Jeffrey Ullman receives ACM Turing Award”, Stanford News, del 31 de marzo de 2021, en <https://news.stanford.edu/2021/03/31/jeffrey-ullman-receives-acm-turing-award/> [Actualizado a julio de 2022]

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Belmont Report, Department of Health, Education, and Welfare, Abril, 1979

Beltrán Jordi Ferrer y Giannini, Leandro J., *Contra la carga de la prueba*, Introducción de Michele Taruffo, Marcial Pons, Madrid, 2019. Consúltese en: <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html> [Actualizado a julio de 2022]

Nadschläger Stefan y Küng, Josef, “Towards a pattern language for knowledge processing systems: expert systems” en *Proceedings of the VikingPLoP 2017 Conference on Pattern Languages of Program (VikingPLoP '17)*, Association for Computing Machinery, New York, 2017, article 1, 1–33. DOI: <https://doi.org/10.1145/3158491.3158506> [Actualizado a julio de 2022]

Nagel, Thomas, *The View from Nowhere*, Oxford, Oxford University Press, 1986. [Existe una traducción al español del Fondo de Cultura Económica (México, 1996)]

National Security Commission on Artificial Intelligence (NSCAI), *The NSCAI Final Report*, 2021. Consúltese en: <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf> [Actualizado a julio de 2022]

Newell, Allen, Shaw John Cliff y Simon Herbert A., “Empirical Explorations of The Logic Theory Machine: A Case Study in Heuristic”, *Computers and Thought* [E.A. Feigenbaum y J Feldman] Mc Graw Hill, New York, 1963. <http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=33559> [Actualizado a julio de 2022]

Nieva-Fenoll, Jordi, *Inteligencia artificial y proceso judicial*, Marcial Pons, Madrid, 2018.

Nieva-Fenoll, Jordi, “Carga de la prueba y estándares de prueba: dos reminiscencias del pasado”, en *InDret* 3. 2020.

Nieva-Fenoll, Jordi, “Los sistemas de valoración de la prueba y la carga de la prueba: Nociones que precisan revisión”, en *Justicia: revista de derecho procesal*, No. 3-4, J. M. Bosch Editor, 2011.

Ningrum DNA, Yuan SP, Kung WM, Wu CC, Tzeng IS, Huang CY, Li JYC, Wang YC, “Deep Learning Classifier with Patient’s Metadata of Dermoscopic Images in Malignant Melanoma Detection”. *J Multidiscip Healthc*, número 14, 2021. Consúltese en: <https://doi.org/10.2147/JMDH.S306284>. [Actualizado a julio de 2022]

Nissenbaum, Helen, “Accountability in a Computerized Society”, en *Science and Engineering Ethics*, número 2, vol. 25, 1996. Consúltese en: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.77> [Actualizado a julio de 2022]

Nubiola, Jaime, *El taller de la Filosofía. Una introducción a la escritura filosófica*, Ediciones Universidad de Navarra, Pamplona, 2017.

Nunan D, O’Sullivan J, Heneghan C, et. al., “Ten essential papers for the practice of evidence-based medicine”, *BMJ Evidence-Based Medicine*, febrero 2018. Consúltese en: <https://ebm.bmj.com/content/22/6/202> [Actualizado a julio de 2022]

OCDE, *Directrices de privacidad de la OCDE*, OCDE, 2013. Consúltese en <http://www.oecd.org/digital/ieconomy/privacy-guidelines.htm> [Actualizado a julio de 2021]

OCDE, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449. Consúltese en: <https://www.oecd.org/going-digital/ai/principles/> [Actualizado a julio de 2022]

Ochigame, Rodrigo. “THE INVENTION OF “ETHICAL AI. How Big Tech Manipulates Academia to Avoid Regulation” en *The Intercept*, 20 de diciembre, 2019.

<https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence> [Actualizado a julio de 2022]

Odell, Steve y Narvil Jasper, “Estudio Nueva tecnología: proyección del impacto económico total (Total Economic Impact)”, *Forrester-IBM*, s/l., febrero de 2020.

Oficina del Alto Comisionado de Derechos Humanos de las Naciones Unidas, *El Derecho Internacional de los Derechos Humanos*, ONU, s/f. <https://www.ohchr.org/sp/professionalinterest/pages/internationallaw.aspx#:~:text=La%20Declaraci%C3%B3n%20Universal%2C%20junto%20con,Carta%20Internacional%20de%20Derechos%20Humanos%E2%80%9D>. [Actualizado a julio de 2022]

ONU, *Pacto Internacional de Derechos Económicos, Sociales y Culturales, Adoptado y abierto a la firma, ratificación y adhesión por la Asamblea General en su resolución 2200 A (XXI)*, 16 de diciembre de 1966. Consúltese en: <https://www.ohchr.org/sp/professionalinterest/pages/cescr.aspx> [Actualizado a julio de 2022]

Organización Mundial de la Salud, *Asamblea Mundial de la Salud, Documentos básicos, suplemento de la 45ª edición*, octubre de 2006, WHO, Ginebra, consúltese en: https://www.who.int/governance/eb/who_constitution_sp.pdf [Actualizado a julio de 2022]

Organización Panamericana de la Salud, *Ocho principios rectores de la transformación digital del sector de la salud. Un llamado a la acción panamericana OPS/EIH/IS/21-2004*, Washington, 2021.

Orrego Puelles, Enrique, “Mal praxis médica”, en *Acta cancerológica*, volumen 31. No. 1, mayo 2002.

Osler, William, *Principles and Practice of Medicine*, Appleton, New York, 1892.

Panel for the Future of Science and Technology, *A governance framework for algorithmic accountability and transparency*, European Parliamentary Research Service of the

Secretariat of the European Parliament. Bruselas, 2019. Consúltese en: [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU\(2019\)624262_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf) [Actualizado a julio de 2022]

Paniagua, Esther, Recommendation of the Council on Artificial Intelligence, Invertia, El Español, 30 de enero de 2021, Consúltese en: https://www.elespanol.com/invertia/disruptores-innovadores/opinion/20210130/lecciones-practicas-inteligencia-artificial-etica/554574546_13.html?utm_term=Autofeed&utm_campaign=Echobox&utm_medium=Social&utm_source=Linkedin [Actualizado a julio de 2022]

Papadopoulos, Leticia. "How Watson AI is helping companies stay ahead of hackers and cybersecurity attacks", IBM, 14 de agosto de 2017, Consúltese en: <https://www.ibm.com/blogs/watson/2017/08/how-watson-ai-is-helping-companies-stay-ahead-of-cybersecurity-attacks/> [Actualizado a julio de 2022]

Parlamento Europeo, Policy Department for Citizens' Rights and Constitutional Affairs, Committee on Legal Affairs, *Artificial Intelligence and Civil Liability, Legal Affairs*, Parlamento Europeo, Bruselas, 2020. Consúltese en: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU\(2020\)621926_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf) [Actualizado a julio de 2022]

Pasquale, Frank, "Machines Judging Humans: The Promise and Perils of Formalizing Evaluative Criteria", en *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*, Association for Computing Machinery, New York, 2020.

-----, *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, Harvard University Press, Massachusetts, 2015. Consúltese en: <https://doi.org/10.4159/harvard.9780674736061> [Actualizado a agosto de 2022]

Patel, N. M., “Enhancing Next-generation sequencing-guided cancer care through cognitive computing”, *Oncologist* número 23, 2018.
<https://doi.org/10.1634/theoncologist.2017-0170>. [Actualizado a julio de 2022]

Paz, Octavio, “El siglo XX: La experiencia de la libertad”, *Vuelta*, número 167, 8 de octubre de 1990.

Picos, Arturo, “Prologo”, Para entender a Carlos Llano, FCE, México, 2014

Piquard, Alexandre, “A Bercy, une cellule d’informaticiens pour aider l’Etat à réguler les GAFA”, en *LeMonde*, 22 de enero de 2021 en
https://www.lemonde.fr/economie/article/2021/01/22/a-bercy-une-cellule-d-informaticiens-pour-aider-l-etat-a-reguler-les-gafa_6067191_3234.html
[Actualizado a julio de 2022]

Ponce de León, Armenta, Luis, *Metodología del derecho*, Porrúa, México, 2020.

Ponce de León C., Ma. Eugenia, “Enseñanza de la medicina basada en la evidencia”, *Revista de la Facultad de Medicina*, UNAM, Vol. 44, número 3, mayo-junio, México, 2001.

Porter, Roy, *The Cambridge Illustrated history of medicine*, Cambridge University Press, Cambridge, 1996.

Powles, Julia, y Chaparro, Enrique, “How Google determined our right to be forgotten”, *The Guardian*, 18 de febrero, 2015.
<https://www.theguardian.com/technology/2015/feb/18/the-right-be-forgotten-google-search> [Actualizado a julio de 2022]

Premios frontera del conocimiento 2016, “La Fundación BBVA premia a Geoffrey Hinton artífice de las primeras máquinas capaces de aprender como el cerebro humano”, Fundación BBVA, Salamanca, 17 de enero 2017. Consúltase en
<https://www.premiosfronterasdelconocimiento.es/noticias/la-fundacion-bbva->

[premia-al-artifice-las-primeras-maquinas-capaces-aprender-cerebro-humano/](#)

[Actualizado a julio de 2022]

Prinsloo, Paul, “Of ‘Black Boxes’ and Algorithmic Decision-Making in (Higher) Education – A Commentary” en *Big Data & Society*, enero, 2020. <https://doi.org/10.1177/2053951720933994>. [Actualizado a julio de 2022]

Ramírez, Fran, “Historia de la IA: Frank Rosenblatt y el Mark I Perceptrón, el primer ordenador fabricado específicamente para crear redes neuronales en 1957”, *LUCA*, 20 de julio de 2018. Consúltese en: <https://web.archive.org/web/20180722124753/https://data-speaks.luca-d3.com/2018/07/historia-de-la-ia-frank-rosenblatt-y-el.html#> [Actualizado a julio de 2022]

Raz, Joseph, “Responsibility and the Negligence”, en *Oxford Journal of Legal Studies*, Volume 30, número 1, Standard, 2010, pp. 1-18. Consúltese en: <https://doi.org/10.1093/ojls/qqq002> [Actualizado a julio de 2021]

Razzaki, S., Baker, A., Perov, Y.N., Middleton, K., Baxter, J., Mullarkey, D., Sangar, D., Taliercio, M., Butt, M., Majeed, A., Do Rosario, A., Mahoney, M., & Johri, S. (2018). *A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis*. *ArXiv*, abs/1806.10698 <https://arxiv.org/pdf/1806.10698.pdf> [Actualizado a julio de 2022]

RedHat, “¿QUÉ ES UNA API? ¿Qué son las API y para qué sirven?”, RedHat, s/f. Consúltese en <https://www.redhat.com/es/topics/api/what-are-application-programming-interfaces> [Actualizado a julio de 2022]

“Reglamento ePrivacy: ¿qué hay que saber?. Digital Guide”, *IONOS*, marzo 3, 2021. <https://www.ionos.mx/digitalguide/paginas-web/derecho-digital/eprivacy-reglamento-sobre-privacidad-electronica-en-la-eu/> [Actualizado a julio de 2022]

Rojina Villegas, Rafael, Compendio de Derecho Civil, XXVI edición, Tomo III, Porrúa, México, 2006.

Rome Call for AI Ethics, El Vaticano, 2020. Consulte el texto en: https://www.romecall.org/wp-content/uploads/2021/02/AI-Rome-Call-x-firma_DEF_DEF_con-firme_.pdf [Actualizado a julio de 2022]

Ross, Casey y Swetlitz, Ike, “IBM to Congress: Watson will transform health care, so keep your hands off our supercomputer”, *STAT*, octubre 4, 2017, en <https://www.statnews.com/2017/10/04/ibm-watson-regulation-fda-congress/> [Actualizado a julio de 2022]

Ross, Casey y Swetlitz, Ike, “IBM’s Watson supercomputer recommended ‘unsafe and incorrect’ cancer treatments, internal documents show”, *STAT*, julio 25, 2018, en <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/> [Actualizado a julio de 2022]

Ross, Casey, “Once billed as a revolution in medicine, IBM’s *Watson Health* is sold off in parts” en *STAT*, 21 de enero de 2022, consúltese en <https://www.statnews.com/2022/01/21/ibm-watson-health-sale-equity/> [Actualizado a julio de 2022]

Roselli, Drew, Matthews, Jeanna y Talagala, Nisha, “Managing Bias in AI” en *Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, 2019. DOI: <https://doi.org/10.1145/3308560.3317590> [Actualizado a julio de 2022]

Rosenblatt, F., “The perceptron: a probabilistic model for information storage and organization in the brain”, *Psychological review*, número 65, 6, 1958. <https://pdfs.semanticscholar.org/5d11/aad09f65431b5d3cb1d85328743c9e53ba96.pdf?ga=2.14359735.1221780561.1614451804-388228147.1609568917>. [Actualizado a julio de 2022]

Roser, Max, Esteban Ortiz-Ospina and Hannah Ritchie, Life Expectancy, octubre 2019, consúltese en: <https://ourworldindata.org/life-expectancy> [Actualizado a julio de 2022]

Rouhiainen, Lasse, *Artificial Intelligence: 101 Things You Must Know Today About Our Future - Updated Edition for Post-Covid-19 World* (English Edition). Edition Kindle, 2020.

Roxin, Claus, *Derecho Penal. Parte General. Fundamentos. La estructura de la teoría del delito*, Editorial Civitas, Madrid, 1997.

Rubio Hancock, Jaime, “El dilema del tranvía: ¿debo sacrificar una vida para salvar cinco?”, *Verne*, 31 de marzo de 2017, en: [¿Cómo funciona Adext AI? Explore la IA y amplifiquen | Centro de ayuda de Adext AI](#)

Rudin C., y Wagstaff, K. L., “Machine Learning for science and society”. *Mach Learn*, número 95, 2014. <https://doi.org/10.1007/s10994-013-5425-9> [Actualizado a julio de 2022]

-----, y Carlson, D.E., “The Secrets of *Machine Learning*: Ten Things You Wish You Had Known Earlier to be More Effective at Data Analysis”, *ArXiv*, *abs/1906.01998*, 2019.

Russell, Stuart y Norving, Peter, *Inteligencia artificial. Un enfoque moderno* (segunda edición), Pearson Educación, Madrid, 2004.

Saleiro, P., Kuester, B., Stevens, A., Anisfeld, A., Hinkson, L., London, J., & Ghani, R., “Aequitas: A Bias and Fairness Audit Toolkit” 2018, *ArXiv*, *abs/1811.05577*. [Aequitas – Center for Data Science and Public Policy \(datasciencepublicpolicy.org\)](#) [Actualizado a julio de 2022]

Sandeep Konam, “Where did IBM go wrong with *Watson Health*?” en Quartz, 02 de marzo de 2022, consúltese en: <https://qz.com/2129025/where-did-ibm-go-wrong-with-watson-health/> [Actualizado a julio de 2022]

Sambasivan, Murali, Esmaeilzadeh, Murali *et. al.*, “Intention to adopt clinical decision support systems in a developing country: effect of physician's perceived professional

autonomy, involvement and belief: a cross-sectional study". *BMC medical informatics and decision making*, número 12, 1, 2012.

Samuel, Arthur L., "Some Studies in *Machine Learning* Using the Game of Checkers", en *IBM Journal of Research and Development*, vol. 3, no. 3, July 1959. Doi: 10.1147/rd.33.0210.

San Agustín, "El pecado procede del libre albedrío", *De la verdadera religión*, XIV. 27, en: https://www.augustinus.it/spagnolo/vera_religione/index2.htm [Actualizado a julio de 2022]

Saña, Heleno, *Breve tratado de Ética. Una introducción a la teoría de la moral*, Almuzara, Madrid, 2009.

Schrittwieser, J., Antonoglou, I., Hubert, T. et al., "Mastering Atari, Go, chess and shogi by planning with a learned model", *Nature*, número 588, 2020. <https://doi.org/10.1038/s41586-020-03051-4> [Actualizado a julio de 2022]

Schwab, Klaus, *La cuarta revolución industrial*, Penguin Random House, México, 2017.

Searle, John, *Mentes, Cerebros y Ciencia*, Ediciones Cátedra, Madrid, 1985.

Shailaja, K., Seetharamulu, B. y Jabbar, M. A., "*Machine Learning* in Healthcare: A Review", *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018. doi: 10.1109/ICECA.2018.8474918.

Sheth, D. & Giger, M. L., "Artificial intelligence in the interpretation of breast cancer on MRI", *J. Magn. Reson. Imaging*, volumen 51, 2019. <https://doi.org/10.1002/jmri.26878> [Actualizado a julio de 2022]

Shortliffe, Edward H., "A rule-based computer program for advising physicians regarding antimicrobial therapy selection" en *Proceedings of the 1974 annual ACM conference - Volume 2 (ACM '74)*, Association for Computing Machinery, New York, 1974. DOI: <https://doi.org/10.1145/1408800.1408906> [Actualizado a julio de 2022]

Shweta, Narkar, Zhang, Yunfeng, et. al., “Intelligent user interfaces for algorithmic transparency in emerging technologies”, en *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion (IUI '19)*, Association for Computing Machinery, New York, 2019. DOI: <https://doi.org/10.1145/3308557.3313125> [Actualizado a julio de 2022]

Shweta Narkar, Yunfeng Zhang, Vera Liao, Dakuo Wang, and Justin Weisz. “Model LineUpper: Supporting Interactive Model Comparison at Multiple Levels for AutoML” en *Proceedings of the ACM IUI 2021*, 2021.

Siebel, Thomas M., *Digital Transformation. Survive and Thrive in an Era of Mass Extinction* Rossetta Books, New York, 2019.

Singh H, Petersen LA, Thomas EJ, Understanding diagnostic errors in medicine: a lesson from aviation, Qual Saf Health Care, 2006 Junio 15, doi: 10.1136/qshc.2005.016444. PMID: 16751463; PMCID: PMC2464840. <https://pubmed.ncbi.nlm.nih.gov/16751463/> [Actualizado a julio de 2022]

Siqueiros Fernández, José Enrique y Hernández Sánchez, José Luis, *Ingeniería, Filosofía y Derecho, Diálogo hacia una perspectiva ecosistémica de la Ética en Inteligencia artificial*, México, s/f. [En prensa, Universidad Panamericana, Facultad de Filosofía, México]

Smith, Steve, 5 Ways The IBM Watson Is Changing Health Care, From Diagnosing Disease To Treating It, diciembre 17, 2015 en [5 Ways The IBM Watson Is Changing Health Care, From Diagnosing Disease To Treating It \(medicaldaily.com\)](#) [Actualizado a julio de 2022]

Soni, S. R., Khunteta y Gupta, M., “A review on intelligent methods used in medicine and life science” en *Proceedings of the International Conference & Workshop on Emerging Trends in Technology (ICWET '11)*, Association for Computing Machinery, New York, 2011. DOI: <https://doi.org/10.1145/1980022.1980173> [Actualizado a julio de 2022]

- Standards Committee of the IEEE Systems, Man, and Cybernetics Society, “IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being”, *IEEE SA Standards Board*, 2020. doi: 10.1109/IEEESTD.2020.9084219 <https://standards.ieee.org/content/ieee-standards/en/standard/7010-2020.html> [Actualizado a julio de 2022]
- Stanford, A Simple Alpha (Go) Zero Tutorial, 29 de diciembre de 2017, Stanford <https://web.stanford.edu/~surag/posts/alphazero.html> [Actualizado a julio de 2022]
- Steiner, Christopher, *Automate This: How Algorithms Came to Rule Our World*, Penguin Publishing Group. Edición de Kindle, 2012.
- Straus, Sharon E., *Medicina basada en la evidencia*, Elsevier Health Sciences, Edición de Kindle, 2019. [Actualizado a julio de 2022]
- Strawson Peter, “Freedom and Resentment” en *Freedom and Resentment, and Other Essays*, Londres, Methuen, 1974. Existe una version en español de editorial Paidós, 1995.
- Strickland, E., “IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care,” en *IEEE Spectrum*, vol. 56, no. 4, abril, 2019, doi: 10.1109/MSPEC.2019.8678513 [Actualizado a julio de 2022]
- Styliani Kleanthous, Tsvi Kuflik, Jahna Otterbacher, Alan Hartman, Casey Dugan, and Veronika Bogina, “Intelligent user interfaces for algorithmic transparency in emerging technologies”, en *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion (IUI '19)*, Association for Computing Machinery, New York, 2019. DOI: <https://doi.org/10.1145/3308557.3313125> [Actualizado a julio de 2022]
- Sucar, Germán, *Metodología de la disertación filosófica y jurídica*, Tirant lo Blanch, Valencia, 2019.

Tamayo y Salmorán, Rolando, *Razonamiento y argumentación jurídica. El paradigma de la racionalidad y la ciencia del derecho*, UNAM, México, 2003.

Tansley, Arthur George, "The use and abuse of vegetational concepts and terms". *Ecology*, no. 16, 1935, pp. 284-307. [Reprinted as part of Trudgill S. 2007. Classics in physical geography revisited, Tansley, A.G. "The use and abuse of vegetational concepts and terms", *Ecology*, número 16, 1935, pp. 284-307] http://www2.agroparistech.fr/geeft/Downloads/Training/TropEcol/Archives_2011/5.1a.pdf [Actualizado a julio de 2022]

Taruffo, Michel, *La prueba de los hechos*, trad. Ferrer, Marcial Pons, Madrid 2002.

Taruffo, Michel, *La prueba*, Marcial Pons, Madrid, 2008.

Tellez Valdés, Julio, *La Informática y el Derecho*, INEGI, México 1983.

-----, *Derecho informático*, Quinta Edición, Mc Graw Hill, México, 2021.

Ten Have, Henk, "Re-evaluating professional autonomy in health care", *Theoretical medicine and bioethics* número 21, vol. 5, 2020.

The Law Commission, Imputed Criminal Intent (Director of Public Prosecutions v. Smith), London, Her Majesty's Stationery Office en: <https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2016/08/LC.-010-IMPUTED-CRIMINAL-INTENT-DIRECTOR-OF-PUBLIC-PROSECUTIONS-v.-SMITH.pdf> [Actualizado a julio de 2022]

The Royal Society, *Machine Learning: the power and promise of computers that learn by example*, The Royal Society, Londres, 2017. Consúltase en: <https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf> [Actualizado a agosto de 2022]

Thompson, Neil, C., et al., "Building the Algorithm Commons: Who Discovered the Algorithms that Underpin Computing in the Modern Enterprise?", *Global Strategy*

Journal, no. 11, 2021, pp. 11-33.
<https://onlinelibrary.wiley.com/doi/epdf/10.1002/gsj.1393> [Actualizado a julio de 2022]

Thoreson, Ryan. (2020). Covid-19 Backlash Targets LGBT People in South Korea Government Should Act to Prevent Discrimination. Human Rights Watch. Cfr. <https://www.hrw.org/news/2020/05/13/covid-19-backlash-targets-lgbt-people-south-korea> [Actualizado al 13 de julio de 2022]

Topol, Eric, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books. Edición de Kindle, 2019.

Trejos Buriticá, Omar, *Lógica de programación*, Ediciones de la U, Bogotá, Medellín, 2017.

Tristan Harris Potential Policy Proposals for Regulation of Social Media and Technology Firms Consúltese más detalles en:
https://www.warner.senate.gov/public/_cache/files/d/3/d32c2f17-cc76-4e11-8aa9-897eb3c90d16/65A7C5D983F899DAAE5AA21F57BAD944.social-media-regulation-proposals.pdf [Actualizado a julio de 2022]

Turing, Alan Mathison, "Computing Machinery and Intelligence", *Mind*, volumen LIX, número 236, octubre 1950. Consúltese en:
<https://doi.org/10.1093/mind/LIX.236.433> [Actualizado a julio de 2022]

Uttley, A. M., "Conditional probability machines and conditioned reflexes", en C. E. Shannon & J. McCarthy (Eds.), *Automata studies*. Princeton University Press, Princeton, 1956.

Van de Poel, Ibo, Lambèr Royakkers, Sjoerd D. Zwart, *Moral Responsibility and the Problem of Many Hands*, Routledge, 2018.

Van de Poel, Ibo, Fahlquist, J. N., Doorn, N., Zwart, S., & Royakkers, L. (2012). The problem of many hands: climate change as an example. *Science and engineering*

ethics, 18(1). <https://doi.org/10.1007/s11948-011-9276-0> [Actualizado a julio de 2022]

Vázquez, Carmen, (editora), *Estándares de prueba y prueba científica. Ensayos de epistemología jurídica*, Marcial Pons, Madrid, 2013.

Vega-de Céniga, M., Allegue-Allegue, N., Bellmunt-Montoya, S., López-Espada, C., Riera-Vázquez, R., Solanich-Valldaura, T., Pardo-Pardo, J., *Medicina basada en la evidencia: concepto y aplicación*, Angiología, 2009, número 61 en: <https://www.elsevier.es/es-revista-angiologia-294-articulo-medicina-basada-evidencia-concepto-aplicacion-S0003317009110040> [Actualizado a julio de 2022]

Véliz, Carissa, *Privacy is Power*, Transworld, Edición de Kindle, 2020.

-----, "Moral zombies: why algorithms are not moral agents", *AI & Soc* 36, 2021.
<https://doi.org/10.1007/s00146-021-01189-x>

Vennam, Sai, "Types of cloud computing", IBM Cloud Learn Hub, 18 de agosto de 2020, en <https://www.ibm.com/cloud/learn/cloud-computing> [La traducción es propia] [Actualizado a julio de 2022]

Villoro Toranzo, Miguel, *Metodología del trabajo jurídico: Técnicas del Seminario de Derecho*, Limusa, México, 1999.

Wachter, Sandra, Mittelstadt, Brent y Russell, Chris, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR" *Harvard Journal of Law & Technology*, volume 31, número 2, 2018. Consúltese en: <https://ssrn.com/abstract=3063289> or <http://dx.doi.org/10.2139/ssrn.3063289> [Actualizado a julio de 2022]

Wang, Dakuo, Wang, Liuping, *et. al.*, "Brilliant AI Doctor" in Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment. In Proceedings of the 2021" en *Conference on Human Factors in Computing Systems (CHI '21)*", Association for

Computing Machinery, article 697, New York. DOI:
<https://doi.org/10.1145/3411764.3445432> [Actualizado a julio de 2022]

Wang, X., *et al.*, “Chest X-ray: hospital-scale chest X-ray database and benchmarks on weakly supervised classification and localization of common thorax diseases”, en *Advances in Computer Vision and Pattern Recognition*, Springer, 2019.

Warren y Brandeis, “The Right to Privacy”, *Harvard Review*, volumen 4, número 5, diciembre 15, 1890. <https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf> [Actualizado a julio de 2022]

Webb, Amy, *The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity*, Hachette Book Group, 2019. [Existe la versión en español de Ediciones Culturales Paidós, México, 2021.]

Weidele, Daniel Karl I., Weisz, Justin D., *et. al.*, “AutoAIViz: opening the black box of automated artificial intelligence with conditional parallel coordinates” en *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020.

Weidele, Daniel Karl I., Justin D Weisz, Erick Oduor, Michael Muller, Josh Andres, Alexander Gray, and Dakuo Wang, “AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates”, en *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020.

Whiting, Kate, “Estas son las 10 principales habilidades laborales del futuro - y el tiempo que lleva aprenderlas”, *Foro Económico Mundial*, 22 de octubre de 2020. Consúltese en: <https://es.weforum.org/agenda/2020/10/estas-son-las-10-principales-habilidades-laborales-del-futuro-y-el-tiempo-que-lleva-aprenderlas/?fbclid=IwAR1vCiQhob-R0ZikoCkHlbuctuiNQULxoHeBNsXNCYZFJ3-gjPNbEKktXWw> [Actualizado a julio de 2022]

Wisconsin Dataset UCI, repositorio de *Machine Learning*:
<https://archive.ics.uci.edu/ml/index.php> [Actualizado a julio de 2022]

World Health Organization, WHA60.29 Health technologies, Sixtieth World Health Assembly, Ginebra, 14 a 23 mayo de 2007, Resolutions and Decisions Annexes, consúltese en: https://www.who.int/healthsystems/WHA60_29.pdf?ua=1 [Actualizado a julio de 2022]

Wu, X., Kumar, V., Ross Quinlan, J. *et al.* (2008). “Top 10 algorithms in data mining”. *Know Inf Syst* 14, 1–37. <https://doi.org/10.1007/s10115-007-0114-2> [Actualizado a julio de 2022]

Xiang, Alice, Reconciling Legal and Technical Approaches to Algorithmic Bias (4 de enero de 2021), *Tennessee Law Review*, volume 88, issue 3, 649 (2021), disponible en SSRN: <https://ssrn.com/abstract=3650635> [Actualizado a julio de 2022]

Yuste, R., Goering, S., Arcas, B. *et. Al.*, “Four ethical priorities for neurotechnologies and AI” en *Nature*, 551, 159–163, 2017. Consúltese también el texto en: <https://doi.org/10.1038/551159a> <https://www.nature.com/articles/551159a.pdf> [Actualizado a julio de 2022]

Zaid, Gabriel, “Heurística”, *Letras Libres*, 7 de marzo de 2013, México. Consúltese en: <https://www.letraslibres.com/mexico-espana/heuristica-0> [Actualizado a agosto de 2021]

Zagal, Héctor, Para entender a Carlos Llano, FCE, México, 2014.

Zarsky, T. Z., “Transparent predictions”, *University of Illinois Law Review*, número 4, 2013, pp. 1503-1570.

Zech, J, Badgeley, M., Liu, M., Costa, A., Titano J., and Oermann, E., “Variable generalization performance of a Deep Learning model to detect pneumonia in chest radiographs: A cross- sectional study”, *PLoS Med.*, 2018.

Zikos, Dimitrios, “A Framework to Design Successful Clinical Decision Support Systems”, en *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '17)*, Association for Computing

Machinery, New York, 2017, pp. 185–188. DOI:
<https://doi.org/10.1145/3056540.3064960> [Actualizado a julio de 2022]

Zwinger, Theodor, *Theatrum praxeos medicae*, Basilea, s/f:
https://archive.org/details/b30550075_0001/page/n7/mode/2up [Actualizado a julio
de 2022]

2.- Legislación

Carta Social Europea de 1961, Turín, 18 de octubre de 1961. Consejo de Europa (Estrasburgo). consúltese en:

<https://www.acnur.org/fileadmin/Documentos/BDL/2002/1934.pdf?file=t3/fileadmin/Documentos/BDL/2002/1934> [Actualizado a julio de 2022]

Código Civil Federal

Código de mejores prácticas de gobierno corporativo del Consejo Coordinador Empresarial, Consúltese en: <https://cce.org.mx/2021/05/10/codigo-de-mejores-practicas-de-gobierno-corporativo/> [Actualizado a julio de 2022]

Código Nacional de Procedimientos Penales

Código Penal Federal

Constitución Política de los Estados Unidos Mexicanos, actualizada al 28 de julio de 2021

Convenio de Oviedo, 1 de diciembre de 1999, consúltese en: <https://archivos.juridicas.unam.mx/www/bjv/libros/5/2290/37.pdf> [Actualizado a julio de 2022]

Convenio para la Protección de las Personas con respecto al Tratamiento Automático de Datos Personales (CETS No. 108) en: <https://www.coe.int/en/web/data-protection/convention108-and-protocol> [Actualizado a julio de 2022]

Decreto Promulgatorio del Convenio para la Protección de las Personas con respecto al Tratamiento Automatizado de Datos de Carácter Personal, Diario Oficial de la Federación del 28 de septiembre de 2018.

EDPB-EDPS Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act), Version 1.1., junio 2021. Consúltese en:

https://edpb.europa.eu/system/files/2021-03/edpb-edps_joint_opinion_dga_en.pdf

[Actualizado a agosto de 2021]

Iniciativa de Ley de Equidad Algorítmica de 2020, consúltase en:
<https://www.govtrack.us/congress/bills/116/s5052/text/is> [Actualizado a julio de 2022]

Iniciativa de Ley de Responsabilidad Algorítmica de 2019, consúltase en:
<https://www.govtrack.us/congress/bills/116/s1108/text> [Actualizado a julio de 2022]

Ley Federal del Trabajo

Ley General de Responsabilidades Administrativas

Ley General de Salud

Norma ISO 31000: 2018

Proyecto de Norma Oficial Mexicana PROY-NOM-241-SSA1-2018, Buenas prácticas de fabricación de dispositivos médicos, DOF, 30 de octubre de 2012. Consúltase en:
https://www.dof.gob.mx/nota_detalle.php?codigo=5562796&fecha=14/06/2019#gs.c.tab=0 [20 julio de 2022]

REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)
<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679&from=ES> [Actualizado a julio de 2022]

Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Consúltase en:

https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683

[Actualizado a julio de 2022]

XLII sesión de la Asamblea Global de Privacidad (octubre de 2020)
<https://globalprivacyassembly.org/document-archive/adopted-resolutions/>

3.- Jurisprudencia

“DAÑOS Y PERJUICIOS. DEBEN SER CONSECUENCIA INMEDIATA Y DIRECTA DEL EVENTO DAÑOSO.” [TA]; 10a. Época; T.C.C.; Gaceta S.J.F.; Libro 60, noviembre de 2018; Tomo III; Pág. 2207 I.8o.C.68 C (10a.) Registro digital: 2018297

“DERECHO A UNA JUSTA INDEMNIZACIÓN POR VIOLACIONES A DERECHOS HUMANOS. SU RELACIÓN CON EL DERECHO DE DAÑOS”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 293 1a. CLXXXIX/2018 (10a.) Registro digital: 2018646.

“DERECHO A UNA JUSTA INDEMNIZACIÓN. DETERMINACIÓN DEL QUÁNTUM EN LOS CASOS EN QUE EL ESTADO ES LA PARTE DEMANDADA.” [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 289 1a. CXCIII/2018 (10a.) Registro digital: 2018643

“DERECHO A UNA JUSTA INDEMNIZACIÓN. PARA DETERMINAR EL MONTO INDEMNIZATORIO, SE DEBE ATENDER A LA MULTIPLICIDAD DE CONSECUENCIAS DEL HECHO ILÍCITO”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 290 1a. CLXXXVII/2018 (10a.) Registro digital: 2018644

“DERECHO A UNA JUSTA INDEMNIZACIÓN. PARÁMETROS QUE RIGEN LA INDIVIDUALIZACIÓN DE LOS MONTOS QUE LA INTEGRAN”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 292 1a. CXC/2018 (10a.) Registro digital: 2018645

“DERECHO FUNDAMENTAL A UNA REPARACIÓN INTEGRAL O JUSTA INDEMNIZACIÓN. SU CONCEPTO Y ALCANCE.”. [J]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 41, abril de 2017; Tomo I; Pág. 752 1a./J. 31/2017 (10a.) Registro: 2014098.

“PRUEBAS EN EL SISTEMA PROCESAL PENAL ACUSATORIO. SU VALORACIÓN LIBRE Y LÓGICA POR EL JUZGADOR EN TÉRMINOS DEL ARTÍCULO 20, APARTADO A, FRACCIÓN II, DE LA CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 69, agosto de 2019; Tomo II; Pág. 1320 1a. LXXIV/2019 (10a.) Registro digital: 2020480

“PRUEBAS EN EL SISTEMA PROCESAL PENAL ACUSATORIO. SU VALORACIÓN LIBRE Y LÓGICA POR EL JUZGADOR EN TÉRMINOS DEL ARTÍCULO 20, APARTADO A, FRACCIÓN II, DE LA CONSTITUCIÓN POLÍTICA DE LOS ESTADOS UNIDOS MEXICANOS. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 69, agosto de 2019; Tomo II; Pág. 1320 1a. LXXIV/2019 (10a.) Registro digital: 2020480

“PRUEBAS. EL OBJETIVO DEL SISTEMA DE VALORACIÓN LIBRE ES EL ESCLARECIMIENTO DE LOS HECHOS SIN NECESARIAMENTE BUSCAR LA VERDAD ABSOLUTA, SINO LA PROBABILIDAD MÁS RAZONABLE.” [TA]; 10a. Época; T.C.C.; Gaceta S.J.F.; Libro 77, agosto de 2020; Tomo VI; Pág. 6214 I.4o.A.44 K (10a.) Registro digital: 2021913

“REPARACIÓN INTEGRAL DEL DAÑO POR VIOLACIONES A DERECHOS HUMANOS. ORIGEN DE SU INCORPORACIÓN AL TEXTO CONSTITUCIONAL EN LA REFORMA DE 10 DE JUNIO DE 2011”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 400 1a. CCCXXXVII/2018 (10a.) Registro digital: 2018805

“REPARACIÓN INTEGRAL DEL DAÑO POR VIOLACIONES A DERECHOS HUMANOS. TOPES MÍNIMOS Y MÁXIMOS DE LA CUANTIFICACIÓN DE LAS

INDEMNIZACIONES, SU INCONSTITUCIONALIDAD”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 402 1a. CXCV/2018 (10a.) Registro digital: 2018806

“RESPONSABILIDAD CIVIL EXTRA CONTRACTUAL EN MATERIA MÉDICO-SANITARIA. SUPUESTOS DE VALORACIÓN EN UN ACTO MÉDICO COMPLEJO.”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 34, septiembre de 2016; Tomo I; Pág. 517 1a. CCXXIX/2016 (10a.) Registro digital: 2012516

“RESPONSABILIDAD CIVIL POR HECHOS PROPIOS, AQUILIANA Y OBJETIVA. DIFERENCIAS.”. [TA]; 9a. Época; T.C.C.; S.J.F. y su Gaceta; Tomo IV, noviembre de 1996; Pág. 512 II.1o.C.T.85 C

“TRANSVERSALIDAD DE LOS DERECHOS HUMANOS. EL QUE UN HECHO PUEDA CALIFICARSE COMO VICTIMIZANTE POR CONLLEVAR VIOLACIONES A AQUÉLLOS, NO IMPLICA HACER A UN LADO LAS REGLAS QUE RIGEN A LAS INSTITUCIONES DISEÑADAS PARA EXIGIR LA REPARACIÓN DEL DAÑO CORRESPONDIENTE, ASÍ COMO A LA DOCTRINA QUE SE HA ELABORADO EN TORNO A ÉSTA”. [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 61, diciembre de 2018; Tomo I; Pág. 464 1a. CLXXXVIII/2018 (10a.) Registro digital: 2018862

RESPONSABILIDAD CIVIL DE LOS HOSPITALES PRIVADOS DERIVADA DE LA NEGLIGENCIA DE SUS MÉDICOS. CARGA DE LA PRUEBA.” [TA]; 10a. Época; 1a. Sala; Gaceta S.J.F.; Libro 16, marzo de 2015; Tomo II; Pág. 1112. [El énfasis es propio]

RESPONSABILIDAD PROFESIONAL MÉDICA. DISTINCIÓN ENTRE ERROR Y MALA PRÁCTICA PARA EFECTOS DE SU ACREDITACIÓN.”. [TA]; 10a. Época; T.C.C.; S.J.F. y su Gaceta; Libro XXV, octubre de 2013; Tomo 3; Pág. 1890 I.4o.A.64 A (10a.) Registro digital: 2004785.

4.- Videografía y sitios de internet especializados en IA

Amazon (AWS) What is AWS? <https://www.youtube.com/watch?v=WXHFqTvfFSw&t=1s>;

[Actualizado a julio de 2022]

Amazon staff, "We are implementing a one-year moratorium on police use of Rekognition"

Policy news and views, 10 de junio de 2020. Consúltase en:

<https://www.aboutamazon.com/news/policy-news-views/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition> [Actualizado a julio de 2022]

Amazon Web Services, *What is AWS?*, Amazon Web Services, publicado el 1 de diciembre de 2019. Consúltase en: https://www.youtube.com/watch?v=a9_D53WsUs

[Actualizado a julio de 2022]

AWS <https://aws.amazon.com/es/sagemaker/clarify> [Actualizado a julio de 2022]

BM Archives Valuable resources on IBM's history, en el sitio:

<https://www.ibm.com/ibm/history/index.html> [Actualizado a julio de 2022]

Carlos Llano sobre el libro "Dilemas éticos de la empresa contemporánea", en el sitio:

https://www.youtube.com/watch?v=i5W_8_pSzsQ [Actualizado a julio de 2022]

Carlos Llano y su relación con San Josemaría Escrivá, fundador del Opus Dei

<https://www.carlosllanocatedra.org/blog-carlos-llano/10-aniversario-muerte-carlos-llano-y-50-visita-san-josemaria-escriva-al-ipade> [Actualizado a julio de 2022]

Centro de Tecnología Cognitiva de Hospitales Ángeles

<https://www.hospitalesangeles.com/saludyvida/articulo.php?id=5763#> [Actualizado a julio de 2022]

Colegio de economistas de Oaxaca, "Analítica e Inteligencia Artificial: El caso de éxito de BANORTE", Conferencia impartida por José Antonio Murillo, director general de Analítica del Grupo Financiero Banorte, 11 de febrero de 2021. Consúltase en:

<https://www.youtube.com/watch?v=QZr2UxIJm7o> [Actualizado a julio de 2022]

DeepMind, “AlphaGo Zero: Discovering new knowledge”, 18 de octubre de 2017.
Consúltese en: <https://www.youtube.com/watch?v=WXHFqTvFfSw&t=1s>
[Actualizado a julio de 2022]

El hombre contra la máquina en <https://www.youtube.com/watch?v=JodijRqCg6k>
[Actualizado al 15 de julio de 2022]

Éticas Research & Consulting <https://www.eticasconsulting.com/> [Actualizado a julio de 2022]

Garry Kasparov versus Deep Thought Documentary” en
<https://www.youtube.com/watch?v=ke8pq-cpOGk> [Actualizado al 15 de julio de 2022]

Google (DeepMind –ALPHA GO ZERO) AlphaGo Zero: Discovering new knowledge
<https://www.youtube.com/watch?v=WXHFqTvFfSw&t=1s> [Actualizado a julio de 2022]

Google, *Exploring 6 AI Myths*. Google, 2020. <https://ai.google/static/documents/exploring-6-myths.pdf> [Actualizado a julio de 2022]

GuiaSalud: www.guiasalud.es [Actualizado a julio de 2022]

Guidelines International Network www.g-i-n.net [Actualizado a julio de 2022]

H2O <https://www.h2o.ai> [Actualizado a julio de 2022]

<https://ai.facebook.com/datasets/casual-conversations-dataset/> [Actualizado a julio de 2022]

https://ai.facebook.com/research?utm_source=hp#fundamental-and-applied [Actualizado a julio de 2022]

<https://ai.google/tools/> [Actualizado a julio de 2022]

<https://aif360.mybluemix.net/> [Actualizado a julio de 2022]

<https://aws.amazon.com/es/> [Actualizado a julio de 2022]

<https://azure.microsoft.com/es-mx/> [Actualizado a julio de 2022]

<https://www.causalens.com/> [Actualizado a julio de 2022]

<https://cloud.google.com/> [Actualizado a julio de 2022]

<https://cloud.ibm.com/> [Actualizado a julio de 2022]

<https://cloud.ibm.com/catalog/services/natural-language-understanding> [Actualizado a julio de 2022]

<https://cloud.ibm.com/catalog/services/natural-language-understanding> [Actualizado a julio de 2022]

<https://cloud.ibm.com/catalog/services/speech-to-text> [Actualizado a julio de 2022]

<https://conferences.computer.org/chase2021/index.html> [Actualizado a julio de 2022]

<https://covid19.who.int/> [Actualizado a julio de 2022]

<https://developer.ibm.com/es> [Actualizado a julio de 2022]

<https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-processing-job-run.html> Aequitas

<https://ebm.jamanetwork.com/> [Actualizado a julio de 2022]

https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683 [Actualizado a julio de 2022]

<https://la.mathworks.com/> [Actualizado a julio de 2022]

<https://psicologiaymente.com/personalidad/5-grandes-rasgos-de-personalidad> [Actualizado a julio de 2022]

<https://uautonoma.cl/wp-content/uploads/2018/10/Di%CC%81ptico-Minor-IAAd.pdf> [Actualizado a julio de 2022]

<https://www.aaai.org/> [Actualizado a julio de 2022]

<https://www.aboutamazon.com/news/policy-news-views/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition> [Actualizado a julio de 2022]

<https://www.carlosllanocatedra.org/> [Actualizado a julio de 2022]

<https://www.cminds.co/reports> [Actualizado a julio de 2022]

<https://www.forbes.com/sites/adigaskell/2018/08/09/integrating-innovation-into-healthcare/?sh=156b0731632e> [Actualizado a julio de 2022]

<https://www.gartner.com/en> [Actualizado a julio de 2022]

<https://www.gob.mx/cedn> [Actualizado a julio de 2022]

<https://www.healthcare.digital/single-post/2020/01/29/Digital-Health-Hype-Cycle-2020>
[Actualizado a julio de 2022]

<https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/>
[Actualizado a julio de 2022]

<https://www.ibm.com/case-studies/abu-dhabi-national-oil-company-adnoc> [Actualizado a julio de 2022]

<https://www.ibm.com/case-studies/woodside-energy-watson-cognitive> [Actualizado a julio de 2022]

<https://www.ibm.com/cloud/compare-and-comply> [Actualizado a julio de 2022]

<https://www.ibm.com/cloud/machine-learning> [Actualizado a julio de 2022]

<https://www.ibm.com/cloud/watson-assistant/> [Actualizado a julio de 2022]

<https://www.ibm.com/cloud/watson-knowledge-catalog> [Actualizado a julio de 2022]

<https://www.ibm.com/cloud/watson-knowledge-catalog> [Actualizado a julio de 2022]

<https://www.ibm.com/cloud/watson-openscale/> [Actualizado a julio de 2022]

<https://www.ibm.com/cloud/watson-studio> [Actualizado a julio de 2022]

<https://www.ibm.com/mx-es/cloud/watson-studio> [Actualizado a julio de 2022]

<https://www.ibm.com/mx-es/watson-health> [Actualizado a julio de 2022]

<https://www.ibm.com/products/maximo/remote-monitoring#demo> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/discovery/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/language-translator/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/natural-language-classifier/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/natural-language-understanding/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/natural-language-understanding/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/personality-insights/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/text-to-speech/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/tone-analyzer/> [Actualizado a julio de 2022]

<https://www.ibm.com/watson/services/visual-recognition/demo/#demo> [Actualizado a julio de 2022]

https://www.ibm.com/watson/stories/bradesco?mhsrc=ibmsearch_a&mhq=Caso%201%20Bradesco [Actualizado a julio de 2022]

https://www.ibm.com/watson/stories/bradesco?mhsrc=ibmsearch_a&mhq=Caso%201%20Bradesco [Actualizado a julio de 2022]

<https://www.ibm.com/watson/stories/ca-en/kone-with-watson.html> [Actualizado a julio de 2022]

<https://www.ibm.com/watson-health/solutions/cancer-research-treatment> [Actualizado a julio de 2022]

https://www.iejcdmx.gob.mx/lista_peritos_auxiliares/ [Actualizado a julio de 2022]

<https://www.iproup.com/innovacion/14477-por-que-microsoft-no-vendera-su-software-de-reconocimiento-facial-a-la-policia> [Actualizado a julio de 2022]

<https://www.mobihealthnews.com/content/forty-five-2016-digital-health-mergers-and-acquisitions> [Actualizado a julio de 2022]

<https://www.onlinedoctor.com/healthcare-jobs-medical-technology/> [Actualizado a julio de 2022]

https://www.sas.com/es_mx/solutions/cloud/sas-cloud.html [Actualizado a julio de 2022]

<https://www.spine-health.com/espanol/lumbalgia/radiculopatia-lumbar> [Actualizado a julio de 2022]

<https://www.tensorflow.org/hub>

<https://www.uba.ar/noticia/20195> [Actualizado a julio de 2022]

IBM (Watson – The debater) The Debater
https://www.youtube.com/watch?v=koec1nnePAM&feature=emb_logo [Actualizado a julio de 2022]

IBM <http://aif360.mybluemix.net> [Actualizado a julio de 2022]

IBM Watson Health en el sitio: <https://www.youtube.com/watch?v=SUR8K-3x7M0&t=94s>
[Actualizado a julio de 2022]

IBM y Fundación Carlos Slim promueven en México el uso de tecnologías para la salud, en <https://fundacioncarlosslim.org/ibm-fundacion-carlos-slim-promueven-en-mexico-uso-tecnologias-la-salud/> [Actualizado a julio de 2022]

IBM, *The debater*, IBM, 11 de marzo de 2020. Consúltese en: https://www.youtube.com/watch?v=koec1nnePAM&feature=emb_logo [Actualizado a julio de 2022]

Llano Cifuentes, Carlos, *Presentación del libro: La enseñanza de la dirección y el método del caso*, en el sitio del IPADE: <https://www.carlosllanocatedra.org/la-ense%C3%B1anza-de-la-direcci%C3%B3n-y-el-m%C3%A9todo-del-caso-rese%C3%B1a> [Actualizado a agosto de 2021]

LI²FE (site123.me) <https://liife.site123.me/> [Actualizado a julio de 2022]

Memorial Sloan Kettering Cancer Center en Nueva York
<https://www.mskcc.org/videos/mskcc-and-ibm-collaborate-applying-watson-technology-help-oncologists> [Actualizado a julio de 2022]

MeSH, Medical Subject Headings www.ncbi.nlm.nih.gov/mesh [Actualizado a julio de 2022]

Microsoft AZURE How does Microsoft Azure work?
<https://www.youtube.com/watch?v=WXHFqTvFsw&t=1s>; [Actualizado a julio de 2022]

National Guidelines Clearinghouse: www.guideline.gov [Actualizado a julio de 2022]

Presentación del libro: La enseñanza de la dirección y el método del caso, en el sitio del IPADE:
<https://www.carlosllanocatedra.org/la-ense%C3%B1anza-de-la-direcci%C3%B3n-y-el-m%C3%A9todo-del-caso-rese%C3%B1a> [Actualizado a julio de 2022]

¿Qué es IBM Watson for Oncology?
<https://centrodetecnologiacognitivahospitalesangeles.com/queeswatson.php>
[Actualizado a julio de 2022]

RED HAT, ¿QUÉ ES UNA API? Qué son las API y para qué sirven, en
<https://www.redhat.com/es/topics/api/what-are-application-programming-interfaces>
[Actualizado a julio de 2022]

Surveillance Nation, BuzzFeed.News, 9 de abril de 2021 en:
<https://www.buzzfeednews.com/article/ryanmac/clearview-ai-local-police-facial-recognition?s=03> [Actualizado a julio de 2022]

Toda la verdad - Vuelo 447 de Air France,
<https://www.youtube.com/watch?v=WC6EGAmO9Wo> [Actualizado a julio de 2022]

The Law Commission, Imputed Criminal Intent (Director of Public Prosecutions v. Smith), London, Her Majesty's Stationery Office en: <https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2016/08/LC.-010-IMPUTED-CRIMINAL-INTENT-DIRECTOR-OF-PUBLIC-PROSECUTIONS-v.-SMITH.pdf> [Actualizado a julio de 2022]

Watson of Oncology: <https://www.youtube.com/watch?v=SUR8K-3x7M0&t=94s>
[Actualizado a julio de 2022]

Whiting, Kate, “Estas son las 10 principales habilidades laborales del futuro - y el tiempo que lleva aprenderlas”, consúltense en:
<https://es.weforum.org/agenda/2020/10/estas-son-las-10-principales-habilidades-laborales-del-futuro-y-el-tiempo-que-lleva-aprenderlas/?fbclid=IwAR1vCiQhob-R0ZikoCkHlbuctuiNQULxoHeBNsXNCYZFJ3-gjPNbEKtXWw> [Actualizado a julio de 2022]

5.- Tesis de doctorado

Cornejo Plaza, María Isabel, *Neurociencia y Derecho. El caso de la mejora cognitiva farmacológica*, Tesis de Doctorado, Universidad de Chile. Repositorio Institucional- Facultad de Derecho, Universidad de Chile, 2021.

Larañaga Monjaraz, Pablo, *El concepto de responsabilidad en la teoría del derecho contemporánea*, Tesis para obtener el grado de doctor en Derecho, director: Manuel Atienza, Universidad de Alicante, Alicante, 1996.

Silvina Bacigalupo Saggese, *La responsabilidad penal de las personas jurídicas. Un estudio sobre el sujeto del Derecho Penal*, Tesis de doctorado, director Dr. D. Gonzalo Rodríguez Mourullo, Universidad Autónoma de Madrid, 1997.

Anexos

Anexo del Capítulo I

Anexo 1

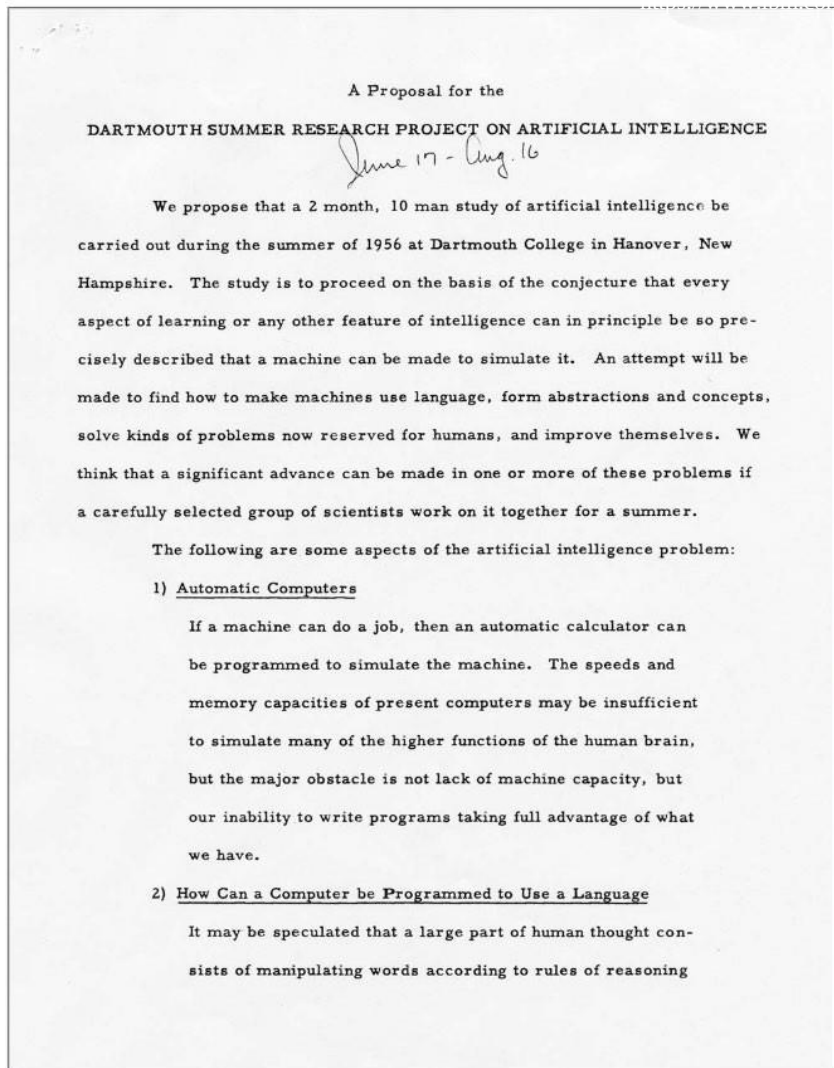


Photo courtesy Dartmouth College.

Page 1 of the Original Proposal.

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

August 31, 1955

*John McCarthy, Marvin L. Minsky,
Nathaniel Rochester,
and Claude E. Shannon*

■ The 1956 Dartmouth summer research project on artificial intelligence was initiated by this August 31, 1955 proposal, authored by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. The original typescript consisted of 17 pages plus a title page. Copies of the typescript are housed in the archives at Dartmouth College and Stanford University. The first 5 papers state the proposal, and the remaining pages give qualifications and interests of the four who proposed the study. In the interest of brevity, this article reproduces only the proposal itself, along with the short autobiographical statements of the proposers.

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use lan-

guage, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1. Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2. How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new

word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out.

3. Neuron Nets

How can a set of (hypothetical) neurons be arranged so as to form concepts. Considerable theoretical and experimental work has been done on this problem by Uttley, Rashevsky and his group, Farley and Clark, Pitts and McCulloch, Minsky, Rochester and Holland, and others. Partial results have been obtained but the problem needs more theoretical work.

4. Theory of the Size of a Calculation

If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order. This method is inefficient, and to exclude it one must have some criterion for efficiency of calculation. Some consideration will show that to get a measure of the efficiency of a calculation it is necessary to have on hand a method of measuring the complexity of calculating devices which in turn can be done if one has a theory of the complexity of functions. Some partial results on this problem have been obtained by Shannon, and also by McCarthy.

5. Self-Improvement

Probably a truly intelligent machine will carry out activities which may best be described as self-improvement. Some schemes for doing this have been proposed and are worth further study. It seems likely that this question can be studied abstractly as well.

6. Abstractions

A number of types of "abstraction" can be distinctly defined and several others less distinctly. A direct attempt to classify these and to describe machine methods of forming abstractions from sensory and other data would seem worthwhile.

7. Randomness and Creativity

A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.

The Proposers

Claude E. Shannon

Claude E. Shannon, Mathematician, Bell Telephone Laboratories. Shannon developed the statistical theory of information, the application of propositional calculus to switching circuits, and has results on the efficient synthesis of switching circuits, the design of machines that learn, cryptography, and the theory of Turing machines. He and J. McCarthy are coediting an *Annals of Mathematics* study on "The Theory of Automata".

Marvin L. Minsky

Marvin L. Minsky, Harvard Junior Fellow in Mathematics and Neurology. Minsky has built a machine for simulating learning by nerve nets and has written a Princeton Ph.D thesis in mathematics entitled, "Neural Nets and the Brain Model Problem" which includes results in learning theory and the theory of random neural nets.

Nathaniel Rochester

Nathaniel Rochester, Manager of Information Research, IBM Corporation, Poughkeepsie, New York. Rochester was concerned with the development of radar for seven years and computing machinery for seven years. He and another engineer were jointly responsible for the design of the IBM Type 701 which is a large scale automatic computer in wide use today. He worked out some of the automatic programming techniques which are in wide use today and has been concerned with problems of how to get machines to do tasks which previously could be done only by people. He has also worked on simulation of nerve nets with particular emphasis on using computers to test theories in neurophysiology.

John McCarthy

John McCarthy, Assistant Professor of Mathematics, Dartmouth College. McCarthy has worked on a number of questions connected with the mathematical nature of the thought process including the theory of Turing machines, the speed of computers, the relation of a brain model to its environment, and the use of languages by machines. Some results of this work are included in the forthcoming "Annals Study" edited by Shannon and McCarthy. McCarthy's other work has been in the field of differential equations.

Anexo 2
Some Studies in *Machine Learning* Using the Game of Checkers
Arthur L. Samuel (1959)

Abstract: Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Further- more, it can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. · The principles of *Machine Learning* verified by these · experiments are, of course, applicable to many other situations.

Introduction: The studies reported here have been concerned with the programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning. While this is not the place to dwell on the importance of *Machine Learning* procedures, or to discourse on the philosophical aspects,¹ there is obviously a very large amount of work, now done by people, which is quite trivial in its demands on the intellect but does, nevertheless, involve some learning. We have at our command computers with adequate data-handling ability and with sufficient computational speed to make use of machine-learning techniques, but our knowledge of the basic principles of these techniques is still rudimentary. Lacking such knowledge, it is necessary to specify methods of problem solution in minute and exact detail, a time-consuming and costly procedure. Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort.

General methods of approach: At the outset it might be well to distinguish sharply between two general approaches to the problem of *Machine Learning*. One method, which might be called the Neural Net Approach, deals with the possibility of inducing learned behavior into a randomly connected switching net (or its simulation on a digital computer) as a result of a reward-and-punishment routine. A second, and much more efficient approach, is to produce the equivalent of a highly organized network which has been designed to learn only certain specific things. The first method should lead to the development of general-purpose learning machines. A comparison between the size of the switching nets that can be reasonably constructed or simulated at the present time and the size of the neural nets used by animals, suggests that we have a long way to go before we obtain practical devices. The second procedure requires reprogramming for each new application, but it is capable of realization at the present time. The experiments to be described here were based on this second approach.

Choice of problem: For some years the writer has devoted his spare time to the subject of *Machine Learning* and has concentrated on the development of learning procedures as applied to games. A game provides a convenient vehicle for such study as contrasted with a problem taken from life, since many of the complications of detail are removed. Checkers, rather than chess, was chosen because the simplicity of its rules permits greater emphasis to be placed on learning techniques. Regardless of the relative merits of the two games as intellectual pastimes, it is fair to state that checkers contains all of the basic characteristics of an intellectual activity in which heuristic procedures and learning processes can play a major role and in which these processes can be evaluated.

Conclusions

As a result of these experiments one can say with some certainty that it is now possible to devise learning schemes which will greatly outperform an average person and that such learning schemes may eventually be economically feasible as applied to real-life problems.

Anexo 3

A Theory of Transfer Learning with Applications to Active Learning

Liu Yang , Steve Hanneke y Jaime Carbonell

Resumen. Exploramos un entorno de aprendizaje por transferencia, en el que una secuencia finita de conceptos objetivo se muestrea de forma independiente con una distribución desconocida de una familia conocida. Estudiamos el número total de ejemplos etiquetados necesarios para aprender todos los objetivos con una precisión esperada especificada arbitrariamente, centrándonos en lo asintótico en el número de tareas y la precisión deseada. Nuestro interés principal es comprender formalmente los beneficios fundamentales del aprendizaje por transferencia, en comparación con el aprendizaje de cada objetivo de forma independiente de los demás. Nuestro enfoque del problema de la transferencia es general, en el sentido de que se puede utilizar con una variedad de protocolos de aprendizaje. Como aplicación particularmente interesante, estudiamos en detalle los beneficios de la transferencia para la autoverificación del aprendizaje activo; en este contexto, encontramos que el número de ejemplos etiquetados necesarios para aprender con la transferencia es a menudo significativamente menor que el necesario para aprender cada objetivo de forma independiente.

Transferir aprendizaje reutiliza el conocimiento de tareas anteriores relacionadas para facilitar el proceso de aprendizaje para realizar una nueva tarea. El objetivo del aprendizaje por transferencia es aprovechar el aprendizaje y la experiencia previos para aprender de manera más eficiente conceptos novedosos pero relacionados, en comparación con lo que sería posible sin esta experiencia previa. La utilidad del aprendizaje por transferencia se mide típicamente por una reducción en el número de ejemplos de entrenamiento requeridos para lograr un desempeño objetivo en una secuencia de problemas de aprendizaje relacionados, en comparación con el número requerido para problemas no relacionados: es decir, menor complejidad de la muestra. En muchos escenarios de la vida real, solo unos pocos ejemplos de capacitación de un nuevo concepto o proceso a menudo son suficientes para que un alumno humano capte el nuevo concepto dado el conocimiento de los relacionados. Por ejemplo, aprender a conducir una furgoneta se convierte en una tarea mucho más fácil si ya hemos aprendido cómo manejar un carro. Aprender francés es algo más fácil si ya hemos aprendido inglés (vs chino), y aprender español es más fácil si sabemos portugués (vs alemán). Por lo tanto, estamos interesados en comprender las condiciones que permiten que una máquina de aprendizaje aproveche el conocimiento abstracto obtenido como subproducto del aprendizaje de conceptos pasados, para mejorar su desempeño en problemas de aprendizaje futuros. Además, nos interesa saber cómo crece la magnitud de estas mejoras a medida que el sistema de aprendizaje adquiere más experiencia al aprender múltiples conceptos relacionados.

El aprendizaje de transferencia está relacionado, al menos en espíritu, con trabajos mucho anteriores sobre aprendizaje analógico y basado en casos [Car83, Car86, VC93, Kol93, Thr96], aunque ese cuerpo de trabajo es anterior al aprendizaje automático moderno y se centra en la reutilización simbólica de problemas pasados, resolviendo soluciones en lugar de en problemas actuales de aprendizaje automático como clasificación, regresión o aprendizaje estructurado. Más recientemente, el aprendizaje por transferencia (y el problema estrechamente relacionado del aprendizaje multitarea) se ha estudiado en casos específicos con enfoques interesantes (aunque a veces heurísticos) [Car97, Sil00, MP04, Bax97, BDS03]. Este artículo considera un marco teórico general para el aprendizaje por transferencia, basado en una perspectiva empírica de Bayes, y deriva resultados teóricos rigurosos sobre los beneficios de la transferencia. A continuación, discutimos la relación de este análisis con el trabajo teórico existente sobre el aprendizaje por transferencia.

1.1 Aprendizaje activo

El aprendizaje activo es una forma poderosa de aprendizaje automático supervisado que se caracteriza por la interacción entre el algoritmo de aprendizaje y la fuente de datos durante el proceso de aprendizaje.

Anexo 4

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN

F. ROSENBLATT

Cornell Aeronautical Laboratory

If we are eventually to understand the capability of higher organisms for perceptual recognition, generalization, recall, and thinking, we must first have answers to three fundamental questions:

1. How is information about the physical world sensed, or detected, by the biological system?
2. In what form is information stored, or remembered?
3. How does information contained in storage, or in memory, influence recognition and behavior?

The first of these questions is in the province of sensory physiology and is the only one for which appreciable understanding has been achieved. This article will be concerned primarily with the second and third questions, which are still subject to a vast amount of speculation, and where the few relevant facts currently supplied by neurophysiology have not yet been integrated into an acceptable theory. With regard to the second question, two alternative positions have been maintained. The first suggests that storage of sensory information is in the form of coded representations or images, with some sort of one-to-one mapping between the sensory stimulus and the stored pattern. According to this hypothesis, if one understood the code or “wiring diagram” of the nervous system, one should, in principle, be able to discover exactly what an organism remembers by reconstructing the original sensory patterns from the “memory traces” which they have left, much as we might develop a photographic negative or translate the pattern of electrical charges in the “memory” of a digital computer. This hypothesis is appealing in its simplicity and ready intelligibility, and a large family of theoretical brain models has been developed around the idea of a coded, representational memory. The alternative approach, which stems from the tradition of British empiricism, hazards the guess that the images of stimuli may never really be recorded at all, and that the central nervous system simply acts as an intricate switching network, where retention takes the form of new connections, or pathways, between centers of activity. In many of the more recent developments of this position (Hebb’s “cell assembly”, and Hull’s “cortical anticipatory goal response”, for example) the “responses” which are associated to stimuli may be entirely contained within the CNS itself. In this case the response represents an “idea” rather than an action. The important feature of this approach is that there is never any simple mapping of the stimulus into memory, according to some code which would permit its later reconstruction.

Whatever information is retained must somehow be stored as a preference for a particular response, i.e., the information is contained in connections or associations rather than topographic representations. (The term response, for the remainder of this presentation, should be understood to mean any distinguishable state of the organism, which may or may not involve externally detectable muscular activity. The activation of some nucleus of cells in the central nervous system, for example, can constitute a response, according to this definition.) Corresponding to these two positions on the method of information retention, there exist two hypotheses with regard to the third question, the manner in which stored information exerts its influence on current activity. The “coded memory theorists” are forced to conclude that recognition of any stimulus involves the matching or systematic comparison of the contents of storage with incoming sensory patterns, in order to determine whether the current stimulus has been seen before, and to determine the appropriate response from the organism. The theorists in the empiricist tradition, on the other hand, have essentially combined the answer to the third question with their answer to the second: since the stored information takes the form of new connections, or transmission channels in the nervous system (or the creation of conditions which are functionally equivalent to new connections), it follows that the new stimuli will make use of these new pathways which have been created, automatically activating the appropriate response without requiring any separate process for their recognition or identification. The theory to be presented here takes the empiricist, or “connectionist” position with regard to these questions. The theory has been developed for a hypothetical

nervous system, or machine called a perceptron. The perceptron is designed to illustrate some of the fundamental properties of intelligent systems in general, without becoming too deeply enmeshed in the special, and frequently unknown, conditions which hold for particular biological organisms. The analogy between the perceptron and biological systems should be readily apparent to the reader. During the last few decades, the development of symbolic logic, digital computers, and switching theory has impressed many theorists with the functional similarity between a neuron and the simple on-off units of which computers are constructed, and has provided the analytical methods necessary for representing highly complex logical functions in terms of such elements. The result has been a profusion of brain models which amount simply to logical contrivances for performing particular algorithms (representing “recall,” stimulus comparison, transformation, and various kinds of analysis) in response to sequences of stimuli —e.g., Rashevsky (14), McCulloch (10), McCulloch & Pitts (11), Culbertson (2), Kleene(8), and Minsky (13). A relatively small number of theorists, like Ashby(1) and von Neumann (17, 18), have been concerned with the problems of how an imperfect neural network, containing many random connections, can be made to perform reliably those functions which might be represented by idealized wiring diagrams. Unfortunately, the language of symbolic logic and Boolean algebra is less well suited for such investigations. The need for a suitable language for the mathematical analysis of events in systems where only the gross organization can be characterized, and the precise structure is unknown, has led the author to formulate the current model in terms of probability theory rather than symbolic logic.

[...]

The position, elaborated by Hebb (7), Hayek (6), Uttley (16), and Ashby (1), in particular, upon which the theory of the perceptron is based, can be summarized by the following assumptions:

1. The physical connections of the nervous system which are involved in learning and recognition are not identical from one organism to another. At birth, the construction of the most important networks is largely random, subject to a minimum number of genetic constraints.
2. The original system of connected cells is capable of a certain amount of plasticity; after a period of neural activity, the probability that a stimulus applied to one set of cells will cause a response in some other set is likely to change, due to some relatively long-lasting changes in the neurons themselves.
3. Through exposure to a large sample of stimuli, those which are most “similar” (in some sense which must be defined in terms of the particular physical system) will tend to form pathways to the same sets of responding cells. Those which are markedly “dissimilar” will tend to develop connections to different sets of responding cells.
4. The application of positive and/or negative reinforcement (or stimuli which serve this function) may facilitate or hinder whatever formation of connections is currently in progress.
5. Similarity, in such a system, is represented at some level of the nervous system by a tendency of similar stimuli to activate the same sets of cells. Similarity is not a necessary attribute of particular formal or geometrical classes of stimuli but depends on the physical organization of the perceiving system, an organization that evolves through interaction with a given environment. The structure of the system, as well as the ecology of the stimulus-environment, will affect, and will largely determine, the classes of “things” into which the perceptual world is divided.

[...]”

The New York Times, 13, July 1958

Electronic 'Brain' Teaches Itself

The Navy last week demonstrated the embryo of an electronic computer named the Perceptron which, when completed in about a year, is expected to be the first non-living mechanism able to "perceive, recognize and identify its surroundings without human training or control." Navy officers demonstrating a preliminary form of the device in Washington said they hesitated to call it a machine because it is so much like a "human being without life."

Dr. Frank Rosenblatt, research psychologist at the Cornell Aeronautical Laboratory, Inc., Buffalo, N. Y., designer of the Perceptron, conducted the demonstration. The machine, he said, would be the first electronic device to think as the human brain. Like humans, Perceptron will make mistakes at first, "but it will grow wiser as it gains experience," he said.

The first Perceptron, to cost about \$100,000, will have about 1,000 electronic "association cells" receiving electrical impulses from an eyelike scanning device with 400 photocells. The human brain has ten billion responsive cells, including 100,000,000 connections with the eye.

Difference Recognized

The concept of the Perceptron was demonstrated on the Weather Bureau's \$2,000,000 IBM 704 computer. In one experiment, the 704 computer was shown 100 squares situated at random either on the left or the right side of a field. In 100 trials, it was able to "say" correctly ninety-seven times whether a square was situated on the right or left. Dr. Rosenblatt said that after having seen only thirty to forty squares the device had learned to

recognize the difference between right and left, almost the way a child learns.

When fully developed, the Perceptron will be designed to remember images and information it has perceived itself, whereas ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons, Dr. Rosenblatt said, will be able to recognize people and call out their names. Printed pages, longhand letters and even speech commands are within its reach. Only one more step of development, a difficult step, he said, is needed for the device to hear speech in one language and instantly translate it to speech or writing in another language.

Self-Reproduction

In principle, Dr. Rosenblatt said, it would be possible to build Perceptrons that could reproduce themselves on an assembly line and which would be "conscious" of their existence.

Perceptron, it was pointed out, needs no "priming." It is not necessary to introduce it to surroundings and circumstances, record the data involved and then store them for future comparison as is the case with present "mechanical brains." It literally teaches itself to recognize objects the first time it encounters them. It uses a camera-eye lens to scan objects or survey situations, and an electrical impulse system, patterned point-by-point after the human brain does the interpreting.

The Navy said it would use the principle to build the first Perceptron "thinking machines" that will be able to read or write.

The New York Times

Published: July 13, 1958

Copyright © The New York Times

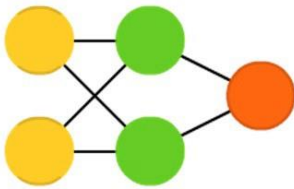
Anexo 6

Van Veen, F. y Leijnen, S. (2019). “The Neural Network Zoo”. Consúltese en <https://www.asimovinstitute.org/neural-network-zoo> [Actualización 22 de abril de 2019]



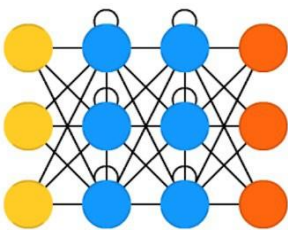
Feedforward redes neuronales (FF o FFNN) y perceptrones (P)

Rosenblatt, Frank. “El perceptrón: un modelo probabilístico para el almacenamiento y la organización de la información en el cerebro”. *Revisión psicológica* 65.6 (1958): 386. [Documento original PDF](#)



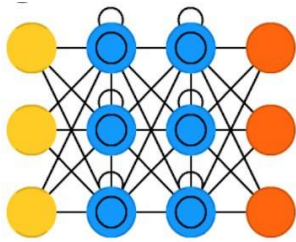
Las redes de función de base radial (RBF)

Broomhead, David S. y David Lowe. *Funciones de base radial, interpolación funcional multivariable y redes adaptativas*. N RSRE-MEMO-4148. ROYAL SIGNALS AND RADAR ESTABLISHMENT MALVERN (REINO UNIDO), 1988. [Documento original PDF](#)



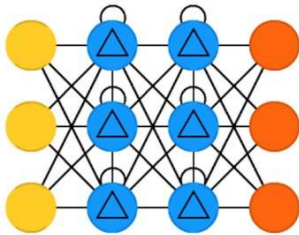
Redes neuronales recurrentes (RNN)

Elman, Jeffrey L. “Encontrar estructura en el tiempo”. *Ciencia cognitiva* 14.2 (1990): 179-211. [PDF en papel original](#)



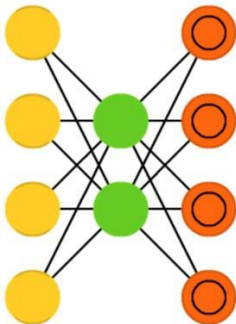
Memoria a largo/corto plazo (LSTM)

Hochreiter, Sepp y Jürgen Schmidhuber. "Memoria a corto plazo". *Computación neuronal* 9.8 (1997): 1735-1780. [PDF en papel original](#)



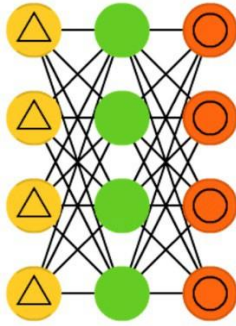
Unidades recurrentes cerradas (GRU)

Chung, Junyoung, et al. "Evaluación empírica de redes neuronales recurrentes cerradas en el modelado de secuencias". preimpresión de *arXiv arXiv*: 1412.3555 (2014). [PDF en papel original](#)



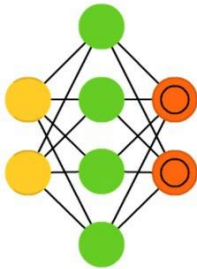
Codificadores automáticos (AE)

Bourlard, Herve e Yves Kamp. "Autoasociación por perceptrones multicapa y descomposición de valores singulares." *Cibernética biológica* 59.4-5 (1988): 291-294. [PDF en papel original](#)



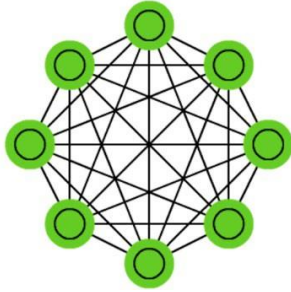
Los codificadores automáticos de eliminación de ruido (DAE)

Vicente, Pascal, et al. "Extracción y composición de características robustas con codificadores automáticos de eliminación de ruido". *Actas de la 25ª conferencia internacional sobre aprendizaje automático. ACM, 2008.* [Documento original PDF](#)



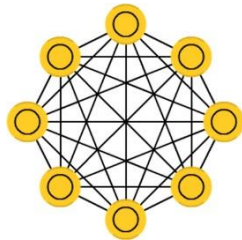
Codificadores automáticos dispersos (SAE)

Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra y Yann LeCun. "Aprendizaje eficiente de representaciones dispersas con un modelo basado en energía". *Procedimientos de NIPS. 2007.* [Documento original PDF](#)



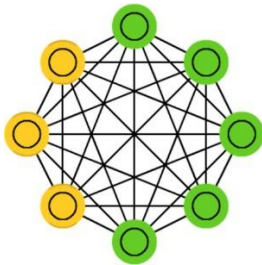
Las cadenas de Markov (MC o cadena de Markov de tiempo discreto, DTMC)

Hayes, Brian. "Primeros eslabones de la cadena de Markov". *American Scientist* 101.2 (2013): 252.
[Documento original PDF](#)



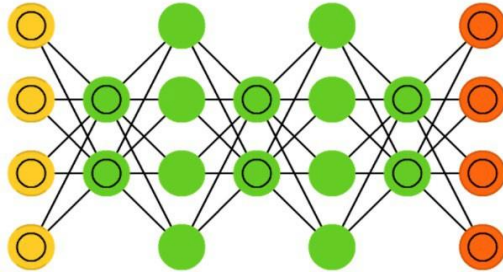
Una red de Hopfield (HN)

Hopfield, John J. "Redes neuronales y sistemas físicos con habilidades computacionales colectivas emergentes". *Actas de la academia nacional de ciencias* 79.8 (1982): 2554-2558 [PDF en papel original](#)



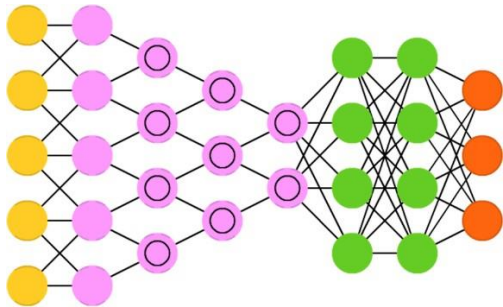
Máquinas Boltzmann (BM)

Hinton, Geoffrey E. y Terrence J. Sejnowski. "Aprendizaje y reaprendizaje en máquinas de Boltzmann". *Procesamiento distribuido en paralelo: exploraciones en la microestructura de la cognición 1* (1986): 282-317
[PDF en papel original](#)



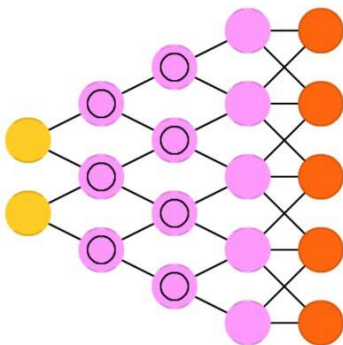
Redes de creencias profundas (DBN)

Bengio, Yoshua, et al. “Entrenamiento codicioso por capas de redes profundas”. *Avances en los sistemas de procesamiento de información neuronal* 19 (2007): 153 [Documento original PDF](#)



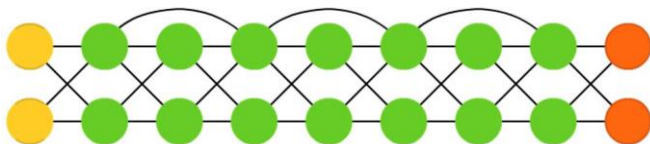
Redes neuronales convolucionales (CNN o redes neuronales convolucionales profundas, DCNN) son bastante diferentes de la mayoría de las otras redes. Se utilizan principalmente para el procesamiento de imágenes, pero también se pueden utilizar para otros tipos de entrada, como audio. Un caso de uso típico para las CNN es cuando alimentas las imágenes de la red y la red clasifica los datos, por ejemplo, genera “gato” si le das una imagen de gato y “perro” cuando le das una imagen de perro.

LeCun, Yann, et al. “Aprendizaje basado en gradientes aplicado al reconocimiento de documentos”. *Actas del IEEE* 86.11 (1998): 2278-2324. [PDF en papel original](#)



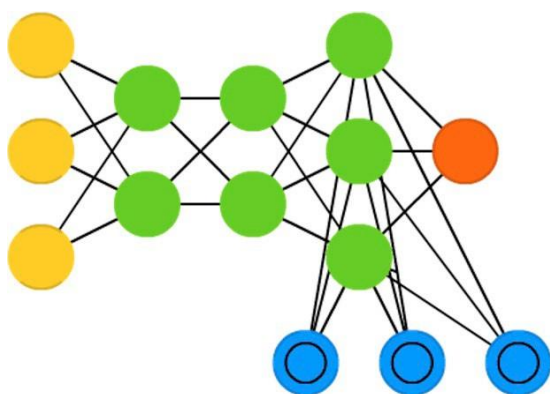
Redes desconvolucionales (DN), también llamadas redes de gráficos inversos (IGN), son redes neuronales convolucionales inversas. Imagine alimentar una red con la palabra “gato” y entrenarla para producir imágenes similares a las de un gato, comparando lo que genera con imágenes reales de gatos.

Zeiler, Matthew D., et al. “Redes desconvolucionales”. *Visión artificial y reconocimiento de patrones (CVPR), Conferencia IEEE de 2010 sobre. IEEE, 2010.* [Documento original PDF](#)

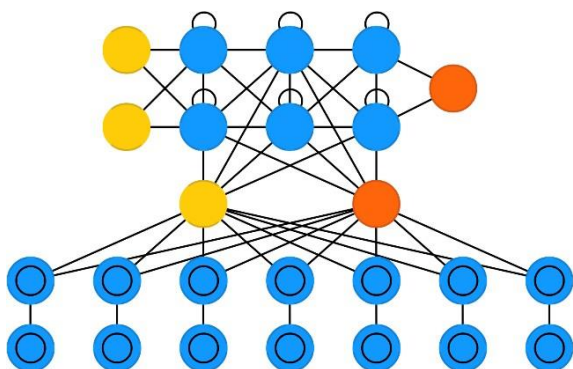


Las redes residuales profundas (DRN)

Él, Kaiming, et al. “Aprendizaje residual profundo para el reconocimiento de imágenes”. *preimpresión de arXiv arXiv:1512.03385 (2015).* [PDF en papel original](#)



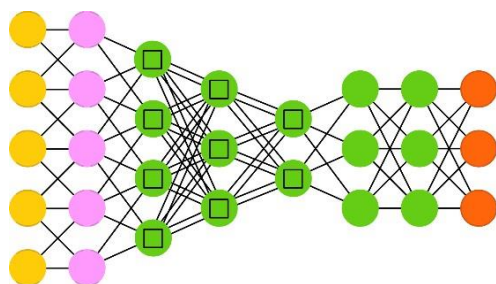
Las máquinas neuronales de Turing (NTM) pueden entenderse como una abstracción de las LSTM y un intento de eliminar las redes neuronales de la caja negra (y darnos una idea de lo que está sucediendo allí). La idea es tener un banco de memoria direccionable por contenido y una red neuronal que pueda leer y escribir desde él. Graves, Alex, Greg Wayne e Ivo Danihelka. “Máquinas de Turing Neurales”. *preimpresión de arXiv arXiv:1410.5401 (2014).* [PDF en papel original](#)



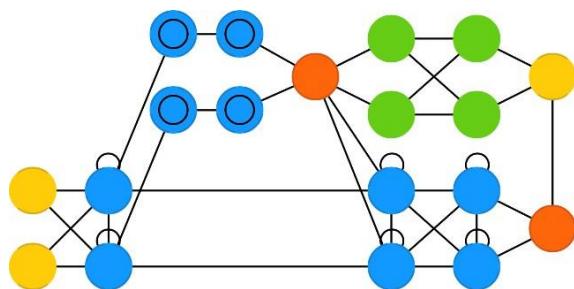
Computadoras neuronales diferenciables (DNC) son máquinas neuronales de Turing mejoradas con memoria escalable, inspiradas en cómo almacena los recuerdos el hipocampo humano. La idea es tomar la arquitectura de computadora clásica de Von Neumann y reemplazar la CPU con un RNN. La DNC también

cuenta con tres mecanismos de atención. Estos mecanismos permiten que la RNN consulte la similitud de un bit de entrada con las entradas de la memoria, la relación temporal entre dos entradas cualesquiera en la memoria y si una entrada de la memoria se actualizó recientemente, lo que hace que sea menos probable que se sobrescriba cuando no hay espacios vacíos. memoria disponible.

Graves, Alex, et al. "Computación híbrida utilizando una red neuronal con memoria externa dinámica". *Naturaleza* 538 (2016): 471-476. [PDF en papel original](#)



Capsule Networks (CapsNet) son alternativas inspiradas en la biología a la agrupación, donde las neuronas están conectadas con múltiples pesos (un vector) en lugar de solo un peso (un escalar). Esto permite que las neuronas transfieran más información que simplemente qué característica se detectó, como dónde se encuentra una característica en la imagen o qué color y orientación tiene. El proceso de aprendizaje implica una forma local de aprendizaje hebbiano que valora las predicciones correctas de salida en la siguiente capa. Sabour, Sara, Frost, Nicholas y Hinton, GE "Enrutamiento dinámico entre cápsulas". *En Avances en sistemas de procesamiento de información neuronal* (2017): 3856-3866. [PDF en papel original](#)



Redes de atención (AN) puede considerarse una clase de redes, que incluye la arquitectura *Transformer*. Utilizan un mecanismo de atención para combatir el deterioro de la información almacenando por separado estados de red anteriores y cambiando la atención entre los estados. Los estados ocultos de cada iteración en las capas de codificación se almacenan en celdas de memoria. Las capas de decodificación están conectadas a las capas de codificación, pero también recibe datos de las celdas de memoria filtrados por un contexto de atención. Este paso de filtrado agrega contexto para las capas de decodificación que enfatizan la importancia de características particulares. La red de atención que produce este contexto se entrena utilizando la señal de error de la salida de la capa de decodificación. Además, el contexto de atención se puede visualizar proporcionando información valiosa sobre qué características de entrada se corresponden con qué características de salida.

Jaderberg, Max, et al. "Redes de transformadores espaciales". *En Avances en sistemas de procesamiento de información neuronal* (2015): 2017-2025. [PDF en papel original](#)

Anexos del Capítulo II

Anexo 1

“Principles for accountable algorithms and a social impact statement for algorithms”

Diakopoulos, Nicholas, et al. *FAT/ML* (2017).

Automated decision making algorithms are now used throughout industry and government, underpinning many processes from dynamic pricing to employment practices to criminal sentencing. Given that such algorithmically informed decisions have the potential for significant societal impact, the goal of this document is to help developers and product managers design and implement algorithmic systems in publicly accountable ways. Accountability in this context includes an obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative social impacts or potential harms.

We begin by outlining five equally important guiding principles that follow from this premise:

Algorithms and the data that drive them are designed and created by people -- There is always a human ultimately responsible for decisions made or informed by an algorithm. “The algorithm did it” is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes.

Responsibility

Make available externally visible avenues of redress for adverse individual or societal effects of an algorithmic decision system, and designate an internal role for the person who is responsible for the timely remedy of such issues.

Explainability

Ensure that algorithmic decisions as well as any data driving those decisions can be explained to end-users and other stakeholders in non-technical terms.

Accuracy

Identify, log, and articulate sources of error and uncertainty throughout the algorithm and its data sources so that expected and worst case implications can be understood and inform mitigation procedures.

Auditability

Enable interested third parties to probe, understand, and review the behavior of the algorithm through disclosure of information that enables monitoring, checking, or criticism, including through provision of detailed documentation, technically suitable APIs, and permissive terms of use.

Fairness

Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics (e.g. race, sex, etc).

We have left some of the terms above purposefully under-specified to allow these principles to be broadly applicable. Applying these principles well should include understanding them within a specific context. We also suggest that these issues be revisited and discussed throughout the design, implementation, and release phases of development. Two important principles for consideration were purposefully left off of this list as they are well-covered elsewhere: privacy and the impact of human experimentation. We encourage you to incorporate those issues into your overall assessment of algorithmic accountability as well.

Social Impact Statement for Algorithms

In order to ensure their adherence to these principles and to publicly commit to associated best practices, we propose that algorithm creators develop a Social Impact Statement using the above principles as a guiding structure. This statement should be revisited and reassessed (at least) three times during the design and development process:

- design stage,
- pre-launch,
- and post-launch.

When the system is launched, the statement should be made public as a form of transparency so that the public has expectations for social impact of the system.

The Social Impact Statement should minimally answer the questions below. Included below are concrete steps that can be taken, and documented as part of the statement, to address these questions. These questions and steps make up an outline of such a social impact statement.

Responsibility

Guiding Questions

- Who is responsible if users are harmed by this product?
- What will the reporting process and process for recourse be?
- Who will have the power to decide on necessary changes to the algorithmic system during design stage, pre-launch, and post-launch?

Initial Steps to Take

- Determine and designate a person who will be responsible for the social impact of the algorithm.
- Make contact information available so that if there are issues it's clear to users how to proceed
- Develop a plan for what to do if the project has unintended consequences. This may be part of a maintenance plan and should involve post-launch monitoring plans.
- Develop a sunset plan for the system to manage algorithm or data risks after the product is no longer in active development.

Explainability

Guiding Questions

- Who are your end-users and stakeholders?
- How much of your system / algorithm can you explain to your users and stakeholders?
- How much of the data sources can you disclose?

Initial Steps to Take

- Have a plan for how decisions will be explained to users and subjects of those decisions. In some cases it may be appropriate to develop an automated explanation for each decision.
- Allow data subjects visibility into the data you store about them and access to a process in order to change it.
- If you are using a machine-learning model:

- consider whether a directly interpretable or explainable model can be used.
- describe the training data including how, when, and why it was collected and sampled.
- describe how and when test data about an individual that is used to make a decision is collected or inferred.
- Disclose the sources of any data used and as much as possible about the specific attributes of the data. Explain how the data was cleaned or otherwise transformed.

Accuracy

Guiding Questions

- What sources of error do you have and how will you mitigate their effect?
- How confident are the decisions output by your algorithmic system?
- What are realistic worst case scenarios in terms of how errors might impact society, individuals, and stakeholders?
- Have you evaluated the provenance and veracity of data and considered alternative data sources?

Initial Steps to Take

- Assess the potential for errors in your system and the resulting potential for harm to users.
- Undertake a sensitivity analysis to assess how uncertainty in the output of the algorithm relates to uncertainty in the inputs.
- Develop a process by which people can correct errors in input data, training data, or in output decisions.
- Perform a validity check by randomly sampling a portion of your data (e.g., input and/or training data) and manually checking its correctness. This check should be performed early in your development process before derived information is used. Report the overall data error rate on this random sample publicly.
- Determine how to communicate the uncertainty / margin of error for each decision.

Auditability

Guiding Questions

- Can you provide for public auditing (i.e. probing, understanding, reviewing of system behavior) or is there sensitive information that would necessitate auditing by a designated 3rd party?
- How will you facilitate public or third-party auditing without opening the system to unwarranted manipulation?

Initial Steps to Take

- Document and make available an API that allows third parties to query the algorithmic system and assess its response.
- Make sure that if data is needed to properly audit your algorithm, such as in the case of a machine-learning algorithm, that sample (e.g., training) data is made available.
- Make sure your terms of service allow the research community to perform automated public audits.

- Have a plan for communication with outside parties that may be interested in auditing your algorithm, such as the research and development community.

Fairness

Guiding Questions

- Are there particular groups which may be advantaged or disadvantaged, in the context in which you are deploying, by the algorithm / system you are building?
- What is the potential damaging effect of uncertainty / errors to different groups?

Initial Steps to Take

- Talk to people who are familiar with the subtle social context in which you are deploying. For example, you should consider whether the following aspects of people's identities will have impacts on their equitable access to and results from your system:
 - Race
 - Sex
 - Gender identity
 - Ability status
 - Socio-economic status
 - Education level
 - Religion
 - Country of origin
- If you are building an automated decision-making tool, you should deploy a fairness-aware data mining algorithm. (See, e.g., the resources gathered at <http://fatml.org>).
- Calculate the error rates and types (e.g., false positives vs. false negatives) for different sub-populations and assess the potential differential impacts.

Authors

Nicholas Diakopoulos, University of Maryland, College Park; Sorelle Friedler, Haverford College; Marcelo Arenas, Pontificia Universidad Catolica de Chile, CL; Solon Barocas, Microsoft Research; Michael Hay, Colgate University; Bill Howe, University of Washington; H. V. Jagadish, University of Michigan; Kris Unsworth, Drexel University; Arnaud Sahuguet, Cornell Tech; Suresh Venkatasubramanian, University of Utah; Christo Wilson, Northeastern University; Cong Yu, Google; Bendert Zevenbergen, University of Oxford

Anexo 2

Principios para la administración responsable de un IA confiable de la

Recomendación del Consejo de Inteligencia Artificial

<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Sección 1: Principios para la administración responsable de una IA confiable

II. RECOMIENDA que los Miembros y no Miembros que se adhieran a esta Recomendación (en adelante, los “Adherentes”) promuevan e implementen los siguientes principios para la administración responsable de la IA confiable, que son relevantes para todas las partes interesadas.

III. PIDE a todos los actores de la IA que promuevan e implementen, de acuerdo con sus respectivos roles, los siguientes Principios para la administración responsable de una IA confiable.

IV. SUBRAYA que los siguientes principios son complementarios y deben considerarse en su conjunto.

1.1. Crecimiento inclusivo, desarrollo sostenible y bienestar

Las partes interesadas deben participar de manera proactiva en la administración responsable de una IA confiable en la búsqueda de resultados beneficiosos para las personas y el planeta, como aumentar las capacidades humanas y mejorar la creatividad, promover la inclusión de poblaciones subrepresentadas, reducir las desigualdades económicas, sociales, de género y de otro tipo, y proteger los entornos naturales. , estimulando así el crecimiento inclusivo, el desarrollo sostenible y el bienestar.

1.2. Valores y equidad centrados en el ser humano

a) Los actores de la IA deben respetar el Estado de Derecho, los Derechos Humanos y los valores democráticos durante todo el ciclo de vida del sistema de IA. Estos incluyen libertad, dignidad y autonomía, privacidad y protección de datos, no discriminación e igualdad, diversidad, equidad, justicia social y derechos laborales reconocidos internacionalmente.

B) Para ello, los actores de la IA deben implementar mecanismos y salvaguardas, como la capacidad de determinación humana, que sean apropiados al contexto y consistentes con el estado del arte.

1.3. Transparencia y explicabilidad

Los agentes de IA deben comprometerse con la transparencia y la divulgación responsable de los sistemas de IA. Para ello, deben proporcionar información significativa, adecuada al contexto y coherente con el estado de la técnica:

I. fomentar una comprensión general de los sistemas de IA,

ii. sensibilizar a las partes interesadas sobre sus interacciones con los sistemas de IA, incluso en el lugar de trabajo,

- iii. Permitir que los afectados por un sistema de IA comprendan el resultado, y
- iv. Permitir a los afectados negativamente por un sistema de IA desafiar su resultado basándose en información simple y fácil de entender sobre los factores y la lógica que sirvió de base para la predicción, recomendación o decisión.

1.4. Robustez, seguridad y protección

a) Los sistemas de IA deben ser robustos, seguros y seguros durante todo su ciclo de vida para que, en condiciones de uso normal, uso previsible o mal uso, u otras condiciones adversas, funcionen de manera adecuada y no presenten riesgos de seguridad irrazonables.

B) Con este fin, los actores de la inteligencia artificial deben garantizar la trazabilidad, incluso en relación con los conjuntos de datos, los procesos y las decisiones tomadas durante el ciclo de vida del sistema de inteligencia artificial, para permitir el análisis de los resultados del sistema de inteligencia artificial y las respuestas a las consultas, de manera adecuada al contexto y coherente con el estado del arte. .

C) Los actores de la IA deben, según sus roles, el contexto y su capacidad para actuar, aplicar un enfoque sistemático de gestión de riesgos a cada fase del ciclo de vida del sistema de IA de forma continua para abordar los riesgos relacionados con los sistemas de IA, incluida la privacidad, la seguridad digital, seguridad y sesgo.

1.5. Responsabilidad

Los actores de la IA deben ser responsables del correcto funcionamiento de los sistemas de IA y del respeto de los principios anteriores, en función de sus funciones, el contexto y en consonancia con el estado de la técnica.

Anexo 3

Ejemplo de política de privacidad (Marzo 2021)

Periódico El País (<https://elpais.com/>)

Con su acuerdo, nosotros y nuestros socios usamos cookies o tecnologías similares para almacenar, acceder y procesar datos personales como su visita en este sitio web. Puede retirar su consentimiento u oponerse al procesamiento de datos basado en intereses legítimos en cualquier momento haciendo clic en “Configuración” o en nuestra Política de Cookies en este sitio web.

Nosotros y nuestros socios hacemos el siguiente tratamiento de datos con su consentimiento y/o nuestro interés legítimo:

Almacenar o acceder a información en un dispositivo, Anuncios y contenido personalizados, medición de anuncios y del contenido, información sobre el público y desarrollo de productos, Compartir datos y perfiles no vinculados a su identidad, Datos de localización geográfica precisa e identificación mediante las características de dispositivos. “Configuración”

¿Para qué finalidades se utiliza mi información y quiénes la utilizan?

Este Sitio Web utiliza cookies propias y de otras entidades, para poder acceder y usar su información para las finalidades que se indican a continuación. Si no está de acuerdo con alguna de estas finalidades, podrá personalizar sus opciones a través de esta pantalla.

Nosotros y las empresas que colaboran con nosotros, tales como anunciantes, operadores publicitarios e intermediarios, usaremos su información obtenida a través de las cookies. Para conocer las empresas colaboradoras que incorporan sus cookies en nuestro sitio web puede acceder a través del botón [Ver nuestros socios](#). Puede configurar sus preferencias de consentimiento por separado para cada uno de los socios mencionados.

Información adicional: Puede conocer la información completa sobre el uso de las cookies, su configuración, origen, finalidades y derechos en nuestra [Política de Cookies](#).

Usted permite el uso de las cookies para las siguientes finalidades:

Almacenar o acceder a información en un dispositivo	Rechazar	Aceptar
Seleccionar anuncios básicos	Rechazar	Aceptar
Crear un perfil publicitario personalizado	Rechazar	Aceptar
Seleccionar anuncios personalizados	Rechazar	Aceptar
Crear un perfil para la personalización de contenidos	Rechazar	Aceptar
Seleccionar contenido personalizado	Rechazar	Aceptar
Medir el rendimiento de los anuncios	Rechazar	Aceptar
Medir el rendimiento del contenido	Rechazar	Aceptar
Utilizar estudios de mercado a fin de generar información sobre el público	Rechazar	Aceptar
Desarrollar y mejorar productos	Rechazar	Aceptar
Compartir datos y perfiles no vinculados a su identidad	Rechazar	Aceptar
Analizar activamente las características del dispositivo para su identificación	Rechazar	Aceptar
Utilizar datos de localización geográfica precisa	Rechazar	Aceptar

Si das tu consentimiento para los fines anteriores, también permites que este sitio web y sus socios operen el procesamiento de los siguientes datos: Colejar y combinar fuentes de datos off line. Garantizar la seguridad, evitar fraudes y depurar errores. Recibir y utilizar para su identificación las características del dispositivo que se envían automáticamente. Servir técnicamente anuncios o contenido, y vincular diferentes dispositivos.

POR TODOS NUESTROS SOCIOS [Ver nuestros socios](#)

PRIVACY MANAGEMENT BY

Rechazar todo Aceptar todo

Anexos del Capítulo III

Anexo 1

Editor de reglas

El editor de reglas proporciona algunas herramientas que le ayudan a definir reglas.

- Diccionario

Añada un diccionario y asígnele un nombre de clase. Las palabras que se encuentren que coincidan con las entradas del diccionario se anotarán automáticamente con la clase asignada.

- Expresión regular

Una expresión regular (regex) es una secuencia de caracteres que define un patrón de búsqueda. Un ejemplo básico sería `[A-Z][a-z]*`, que busca palabras que empiezan por mayúscula.

`[A-Z]` coincide con cualquier letra alfabética en mayúscula (de la A a la Z) y `[a-z]*` coincide con cualquier letra alfabética en minúscula (de la a a la z) cero o más veces. El asterisco (*) es el carácter que define el valor de repetición (cero o más veces).

Considere la posibilidad de utilizar un programa de utilidad regex basado en web gratuito para ayudarle a determinar la expresión correcta que se utilizará para capturar el patrón que desea encontrar.

Por ejemplo, sus documentos pueden tener muchas referencias similares a las siguientes frases:

35-year-old driver

16-year-old learner

La sintaxis `n-year-old x` es un patrón que normalmente representa a una persona. Puede definir una regla de expresión regular para buscar frases que coincidan con el patrón `n-year-old x`, y anotarlas como menciones de entidad `PERSON`.

Métodos recomendados para expresiones regulares

Aunque la herramienta regex que se incluye en el editor de reglas reconoce expresiones que siguen la sintaxis de `java.util.regex.Pattern`, no se admite la sintaxis completa. Al utilizar la herramienta regex, tenga en cuenta los siguientes métodos recomendados:

- Las reglas coinciden sólo en los límites de señales (tokens). No se pueden escribir reglas que coincidan con una subsecuencia de una señal. Para obtener más información, consulte [Señaladores](#).
- Sus expresiones regulares deben ser simples y limpias.
- Plántese el uso de diccionarios en lugar de expresiones regulares. Aunque se le ocurra una regla inteligente que puede capturar varias expresiones, en general, la coincidencia de diccionario es más rápida que la coincidencia de reglas. Además, los diccionarios son más fáciles de mantener.
- En lugar de depender completamente en expresiones regulares para la coincidencia de señales (tokens), es mejor utilizar una combinación de diccionarios, expresiones regulares y reglas. Por ejemplo, pensemos en una situación de un número de teléfono en una frase `My móvil es 123-456`. Es posible que pueda escribir una regla que utilice una expresión regular para que coincida con dicha frase. Pero en este caso, el método recomendado es añadir un diccionario para encontrar palabras como *móvil* y

teléfono, escribir una simple expresión regular que capture posibles secuencias de número de teléfono y, a continuación, crear una regla para buscar una secuencia de patrones, como término del diccionario + texto + regex, según se muestra en la frase de ejemplo, Mi móvil es 123-456.

- Evite el uso innecesario de *lookahead* y *lookbehind* (`(?=ABC)`). En muchos casos, puede lograr el mismo resultado utilizando una combinación de expresiones regulares y reglas.

Anexo 2

Adición de diccionarios a un espacio de trabajo

La adición de diccionarios es un paso opcional al crear un modelo. Los diccionarios son útiles porque permiten iniciar rápidamente el proceso de anotación.

Si proporciona un diccionario, puede ejecutar el preanotador de diccionarios en los documentos. El preanotador busca términos representados en el diccionario y los anota automáticamente. Este paso inicial en los documentos simplifica el trabajo del anotador humano porque podrá revisar las anotaciones añadidas por el preanotador y corregirlas o añadirlas a los mismos. No tiene que empezar todo desde cero.

Se aplican las siguientes restricciones a los diccionarios:

- Máximo de 15.000 entradas por diccionario

Nota: Este límite no se aplica a los diccionarios cargados como un archivo  de diccionario. Los diccionarios de solo lectura pueden contener más entradas.

- Máximo de 64 diccionarios por espacio de trabajo

Para añadir un diccionario al espacio de trabajo:

1. Inicie sesión como un administrador o gestor de proyectos de Knowledge Studio y abra la página **Activos > Diccionarios**.
2. Efectúe una de las tareas siguientes:
 - Junto al botón **Crear diccionario**, pulse el icono **Menú** y, a continuación, seleccione **Cargar diccionario**. Seleccione un diccionario y, a continuación, pulse **Cargar**. Después de cargar un diccionario, selecciónelo para ver el diccionario y asociarlo con un tipo de entidad.

Puede cargar un archivo ZIP que contenga un diccionario descargado de otro espacio de trabajo de Knowledge Studio. Debe cargar el sistema de tipos descargado desde el otro espacio de trabajo en formato JSON para poder cargar el archivo de diccionario correspondiente. Puede editar y añadir entradas a un diccionario que reutilice de otro espacio de trabajo de Knowledge Studio. Consulte [Carga de recursos desde otro espacio de trabajo](#) para obtener más detalles.

La carga de un archivo CSV también está soportada, pero cargar directamente como un diccionario crea un diccionario de solo vista previa que no puede editar ni descargar. Para cargar un archivo CSV que pueda editar y descargar, pulse **Crear diccionario** para crear un diccionario en primer lugar y, a continuación, cargar el contenido de CSV como entradas a dicho diccionario creado recientemente.

- Pulse el botón **Crear diccionario** para crear un diccionario vacío al que puede añadir entradas de diccionario. Especifique un nombre descriptivo para el diccionario y, a continuación, pulse **Guardar**.
- 3. Para añadir entradas al diccionario, realice una de las siguientes tareas:
 - Pulse **Añadir entrada** para añadir una entrada de diccionario. Especifique el *lema* (el formato de palabra más representativo para el término).
 - Pulse **Cargar** para cargar un archivo  que contenga entradas de diccionario, y, a continuación, examine para seleccionar el archivo. El archivo  debe ser inferior a 1 MB.
 - Si ha cargado documentos en el espacio de trabajo, seleccione una entrada de diccionario sugerida desde el panel Sugerencias de entrada. Knowledge Studio utiliza sus documentos para determinar las entradas de diccionario probables que se parecen al dominio de los documentos del espacio de trabajo.

4. Después de cargar o de añadir entradas, podrá editar las entradas.

Abra una entrada para especificar términos equivalentes, llamados *formas superficiales*. Cada forma superficial debe tener 256 o menos caracteres. Puede cambiar cuál de las formas superficiales se utilizará como lema. Por ejemplo, el lema *IBM* puede tener formas superficiales como *IBM Corp.* e *International Business Machines, Inc.*.

5. Seleccione la categoría léxica apropiada para cada lema y forma superficial del diccionario.

La información de categoría léxica la utiliza el señalizador, y durante la preanotación.

6. Pulse **Guardar** para almacenar los cambios.

Ejecute el preanotador, que utiliza los diccionarios creados para realizar un pase preliminar de los documentos de origen, y que añade anotaciones a los mismos.

Anexo 3

Glosario hecho por IBM

- **Acuerdo entre anotadores:** Una medida de con cuánta similaridad se anota un documento en dos o más conjuntos de documentos.
- **Adjudicación:** Un proceso iterativo para resolver los conflictos de anotaciones al comparar las anotaciones añadidas al mismo documento por anotadores humanos distintos.
- **Análisis de exactitud:** Analizar el modelo de aprendizaje automático puntuará para determinar si los cambios son necesarios para mejorar la exactitud.
- **Análisis de margen de mejora:** Proceso para determinar cuánta mejora en la exactitud, la precisión o la exhaustividad puede esperarse abordando algunas clases de problemas identificados al realizar un análisis de exactitud.
- **Anotación:** Información sobre un fragmento de texto. Por ejemplo, una anotación podría indicar que un fragmento de texto representa un nombre de empresa.
- **Anotador de aprendizaje automático:** Véase modelo de aprendizaje automático.
- **Anotador humano:** Un experto en la materia que revisa, modifica y aumenta los resultados de la preanotación identificando menciones, relaciones de tipos de entidades, y correferencias de menciones. Al examinar texto en contexto, un anotador humano ayuda a determinar los datos de campo y a mejorar la exactitud del modelo de aprendizaje automático.
- **Anotador:** Consulte anotador humano y anotador de aprendizaje automático.
- **Aprendizaje automático:** Un método de análisis de datos que aprende de forma iterativa a partir de los datos pasados y que se adapta independientemente cuando se expone a datos nuevos. El modelo matemático en el centro de aprendizaje automático se basa en entradas de datos de campo. Mediante el entrenamiento y el refinamiento de los datos de entrada de ejemplo, el modelo puede ofrecer datos precisos y repetibles cuando analiza datos nuevos.
- **Archivador de motor de proceso (PEAR) :** Un archivo de archivado `.pear` que incluye un motor de análisis de UIMA (Unstructured Information Management Architecture) y todos los recursos necesarios para utilizarlo para el análisis personalizado.
- **Atributo:** Característica o rasgo de una entidad que describe la entidad; por ejemplo, el número de teléfono de un empleado es uno de los atributos de empleado.
- **Cadena de correferencia:** Una lista de entidades que se han anotado como correferencias. Al anotar menciones como correferencias, el sistema creará una cadena de correferencia. La cadena proporciona una forma de ver todas las menciones en contexto y de verificar que todas las ocurrencias pertenezcan al mismo tipo de entidad.
- **Característica:** Un miembro de datos o atributo de un tipo.
- **Categoría léxica (POS) :** En un diccionario, se asigna a elementos léxicos individuales etiquetas de categoría léxica (POS). Por ejemplo, la palabra 'fly' se puede identificar como un verbo o un nombre.

- **Concordancia:** Proporciona una forma de asegurar que la misma mención se anote con el mismo tipo de entidad en un documento y en conjuntos de anotaciones. La concordancia ayuda a garantizar la coherencia entre varias apariciones de una mención sin necesitar que el anotador humano etiquete manualmente cada aparición.
- **Conjunto de anotaciones:** En la anotación humana, una recopilación de documentos extraídos desde el corpus que permiten que varios anotadores humanos compartan la carga de trabajo. En la anotación basada en máquina (automática), una recopilación de documentos que se pueden utilizar como datos ciegos, datos de entrenamiento, o datos de prueba.
- **Conjunto de documentos:** Una recopilación de documentos. Los documentos importados juntos se convierten en un conjunto de documentos. Los documentos anotados que se agrupan juntos para fines de entrenamiento (Prueba, entrenamiento, ciegos) se generan como conjuntos de documentos.
- **Conjunto de reglas:** Un conjunto de reglas que definen patrones para anotar texto. Si se aplica un patrón, las acciones de la regla se realizarán en las anotaciones coincidentes. Una regla normalmente especifica la condición que debe coincidir, un cuantificador opcional, una lista de restricciones adicionales que el texto coincidente debe cumplir, y las acciones que se realizarán cuando se produzca una coincidencia, como crear una nueva anotación o modificar una anotación existente.
- **Corpus:** Una recopilación de documentos de origen que se han añadido a un espacio de trabajo y que se han utilizado para entrenar un modelo de aprendizaje automático.
- **Correferencia:** Una relación entre dos palabras o frases en la que ambas hacen referencia a la misma persona o cosa y una se erige como un antecedente lingüístico de la otra. Por ejemplo, hay una correferencia entre los dos pronombres en la frase “She taught herself” (“Se enseñó a sí misma”) pero no en la frase “She taught her” (“Le enseñó a ella”). Una correferencia enlaza dos entidades equivalentes en el mismo texto.
- **Datos ciegos:** Un conjunto de documentos anotados con los datos de campo, como pares de preguntas y respuestas, anotación semántica y decisiones de pasaje. Los datos ciegos nunca los lanzan ni los ven los desarrolladores y se utilizan para probar el sistema periódicamente para evaluar el rendimiento en datos no mostrados. La prueba en datos ciegos impide que se corrompa la exactitud al sobreajustar a conjuntos o anotaciones de preguntas conocidas. Los resultados notificados solo deberían provenir de las pruebas que se ejecutan en los datos ciegos. Véase también [datos de prueba](#) y [datos de entrenamiento](#).
- **Datos de campo:** Conjunto de datos evaluados, que constan de anotaciones añadidas por anotadores humanos, que se utilizan para adaptar un modelo de aprendizaje automático a un dominio particular. Los datos de campo se utilizan para entrenar los modelos de aprendizaje automático, para medir el rendimiento de modelo (precisión y la exhaustividad), y para calcular el margen dinámico para decidir dónde centrar los esfuerzos de desarrollo para mejorar el rendimiento. La exactitud de los datos de campo es esencial, ya que las inexactitudes en los datos de campo se correlacionarán con las inexactitudes en los componentes que los utilizan.
- **Datos de entrenamiento:** Un conjunto de documentos anotados que pueden utilizarse para entrenar modelos de aprendizaje automático. Véase también [datos ciegos](#) y [datos de prueba](#).
- **Datos de prueba:** Un conjunto de documentos anotados que pueden utilizarse para evaluar las métricas de sistema después de la ingesta y del entrenamiento. Véase también [datos ciegos](#) y [datos de entrenamiento](#).
- **Diccionario:** Una recopilación de palabras que pueden utilizarse para preanotar documentos. Se creará una anotación nueva para cada palabra en el texto del documento que coincida con un término del diccionario. Un modelo de aprendizaje automático se puede configurar con uno o varios diccionarios

independientes, que son normalmente específicos del dominio, como un diccionario para productos farmacéuticos y un diccionario para la gestión de la riqueza. Véase también [lema](#) y [forma superficial](#).

En un modelo de relación basada en entidades, un tipo de entidad es lo que se modela o a lo que hace referencia una mención, como por ejemplo el nombre de una persona o de un lugar. Distintos tipos de entidades tienen distintos conjuntos de atributos como “surname” (“apellido”) o “home town” (“ciudad de procedencia”), y están conectados mediante relaciones como “lives in” (“vive en”). Un tipo de entidad existe independientemente y puede identificarse de forma exclusiva.

- **Entidad con nombre:** Un concepto de un dominio que entra en una categoría bien definida, como los nombres de organizaciones, las ubicaciones, los autores o las enfermedades.
- **Entidad:** Mención anotada por un tipo de entidad. Persona, objeto o concepto sobre el que se almacena información. Conjunto de detalles que se mantienen sobre un objeto del mundo real como una persona, una ubicación o una cuenta bancaria. Una entidad es un tipo de elemento.
- **Entrenar:** Proceso de configurar una instancia de Watson con componentes que permiten que el sistema funcione en un dominio determinado (por ejemplo: contenido del corpus, datos de entrenamiento que generan modelos de aprendizaje automático, algoritmos mediante programación u otros componentes de datos de campo) y luego de realizar mejoras y actualizaciones en estos componentes basándose en el análisis de exactitud.
- **Exactitud:** Una medida de la corrección de anotaciones producidas por un modelo de aprendizaje automático. Consulte también [precisión](#) y [exhaustividad](#).
- **Exhaustividad:** Una medida que especifica el porcentaje de resultados relevantes devueltos, de todos los resultados relevantes disponibles. La exhaustividad, que es una medida de sensibilidad, está determinada por el número de resultados positivos correctos dividido por el número de resultados positivos que deberían haberse devuelto. La exactitud se mide mejor mediante la precisión y la exhaustividad. Véase también [exactitud](#) y [precisión](#).
- **Falso negativo:** Una respuesta o una anotación que es correcta, pero que se ha predicho que es incorrecta.
- **Falso positivo:** Una respuesta o anotación que es incorrecta, pero que se ha predicho que es correcta.
- **Forma superficial:** Formato de una palabra o unidad de varias palabras como se encuentra en el corpus. Por ejemplo, algunas formas superficiales del lema 'organize' ('organizar') son los términos 'organizing' ('organización') y 'organized' ('organizado'). Véase también [diccionario](#) y [lema](#).
- **Gestor del proceso de anotación:** Un rol que es responsable de gestionar las actividades de ciclo de vida de anotación completa dentro de un espacio de trabajo. El gestor de proyectos que se añade a un espacio de trabajo suele realizar las actividades de un gestor de procesos de anotación.
- **Gráfico de conocimiento:** Un modelo que consolida las entidades con tipos, sus relaciones, sus propiedades y las taxonomías jerárquicas para representar una organización de conceptos para un dominio determinado. Una vez que el almacén de gráficos de conocimiento se cargue con entradas desde orígenes de datos estructurados y no estructurados, los usuarios y las aplicaciones podrán acceder al gráfico de conocimiento para explorar elementos clave de conocimiento para un dominio específico, para explorar interacciones y para descubrir relaciones adicionales.
- **Lema:** Formato normalizado o canónico de una palabra. Normalmente, el lema es el formato no derivado y no declinado de un nombre o de un verbo. Por ejemplo, el lema de los términos 'organizing' ('organización') y 'organized' ('organizado') es 'organize' ('organizar'). Consulte también [diccionario](#) y [forma superficial](#).

- **Matriz de confusión:** Una tabla que proporciona un desglose numérico detallado de conjuntos de documentos anotados. La tabla se utiliza para comparar las anotaciones añadidas por un modelo de aprendizaje automático a las anotaciones en los datos de campo. La tabla informa del número de falsos positivos, falsos negativos, verdaderos positivos y verdaderos negativos.
- **Mención:** Una parte de texto que considere relevante en los datos de dominio. Por ejemplo, en un sistema de tipos sobre vehículos automotores, las apariciones de términos como “airbag”, “Ford Explorer” y “sistema de retención infantil” podrían ser menciones relevantes.
- **Modelo de aprendizaje automático:** Un componente que identifica entidades y relaciones de entidades según un modelo estadístico que se basa en datos de campo. El modelo se aplica a la experiencia pasada, como los datos de entrenamiento, para determinar o predecir el resultado correcto de experiencias futuras basadas en las características de los datos. Estas experiencias pasadas son captadas en forma de modelo mediante el cálculo de puntuaciones de características para cada prueba o respuesta de un candidato y su combinación con resultados conocidos. A veces conocidas como *anotador de aprendizaje automático*.
- **Motor de análisis:** Un programa que analiza los artefactos, como los documentos, e infiere información sobre ellos, y que implementa la especificación de interfaz del motor de análisis de UIMA. Los motores de análisis se crean desde bloques de construcción llamados anotadores. Un motor de análisis puede contener un único anotador, que se conoce como motor de análisis primitivo, o varios anotadores, que se conocen como un motor de análisis agregado.
- **Ontología:** Especificación formal explícita de la representación de los objetos, conceptos y otras entidades que pueden existir en un área de interés y las relaciones entre ellos.
- **Preanotador de diccionarios:** Un componente que identifica menciones en el texto que coinciden con un conjunto específico de palabras. Al utilizar terminología específica del dominio para preanotar texto, los preanotadores de diccionarios pueden acelerar la capacidad de un anotador humano para preparar un conjunto de documentos de datos de campo.
- **Precisión:** Una medida que especifica la proporción de resultados que son relevantes. La precisión, que es un valor predictivo positivo, se determina mediante el número de resultados positivos correctos dividido por el número de todos los resultados positivos. La exactitud se mide mejor mediante la precisión y la exhaustividad. Véase también exactitud y exhaustividad.
- **Preseleccionar:** Seleccionar, recopilar, conservar y mantener contenido relevante para un tema específico. La preselección establece, mantiene y añade valor a los datos; transforma datos en información y conocimiento de confianza.
- **Procesamiento del lenguaje natural:** Un campo de inteligencia artificial y de lingüística que estudia los problemas inherentes en el proceso y en la manipulación de lenguaje natural, con el fin de aumentar la capacidad de los ordenadores para comprender los idiomas humanos.

Proceso de anotación de un conjunto de documentos antes de la anotación humana. Los documentos se pueden preanotar utilizando un modelo basado en reglas, un modelo de aprendizaje automático, IBM Watson Natural Language Understanding, o un diccionario. La preanotación puede ayudar a los anotadores humanos a preparar más rápidamente un conjunto de documentos de datos de campo.

- **Puntuación de F1:** Una medida de la exactitud de una prueba que considera tanto la precisión como la exhaustividad para calcular la puntuación. La puntuación de F1 se puede interpretar como un promedio ponderado de los valores de precisión y de exhaustividad. Una puntuación de F1 llega a su mejor valor en 1 y a su peor valor en 0.

- **Puntuación de Fleiss Kappa:** Una medida que muestra con cuánta coherencia han aplicado la misma anotación varios anotadores humanos en documentos que se solapan. La puntuación de Fleiss Kappa llega a su mejor valor en 1 y a su peor valor en 0.
- **Relación:** Normalmente un verbo que refleja cómo se relacionan las entidades entre sí. Por ejemplo, “lives in” (“vive en”) es una relación entre una persona y una ciudad. Una relación enlaza dos entidades distintas en la misma frase.
- **Rendimiento:** Medida de un sistema de Watson en términos de exactitud, precisión y exhaustividad, por ejemplo, al responder preguntas, descubrir relaciones o anotar texto.
- **Rol:** Un atributo que proporciona un significado que distingue el contexto de una mención. Por ejemplo, en la frase “I went to IBM today” (“Hoy he ido a IBM”), IBM es la mención, Organization es el tipo de entidad, y Facility es el rol del tipo de entidad.
- **Sistema de tipos:** El sistema de tipos define los tipos de objetos que se pueden descubrir en un documento. El sistema de tipos define todos los tipos de entidades y relaciones posibles entre tipos de entidades. Puede definir cualquier número de tipos distintos en un sistema de tipos. Los sistemas de tipos son generalmente específicos del dominio y de la aplicación.
- **Subtipo:** Tipo que amplía o implementa otro tipo; el supertipo.
- **Tipo de entidad:** El tipo de entidad que representa una mención sin tener en cuenta el contexto. Por ejemplo, la mención IBM podría estar anotada por el tipo de entidad ORGANIZATION.
- **Tipo de relación:** Una relación binaria y unidireccional entre dos entidades. Por ejemplo, Mary employedby IBM es una relación válida; IBM employedby Mary, no.
- **Verdadero negativo:** Una respuesta o anotación que en realidad es incorrecta y se prevé que sea incorrecta.
- **Verdadero positivo:** Una respuesta o anotación que en realidad es correcta y que se prevé que sea correcta.

Creación de diccionarios

Los diccionarios ayudan a los modelos de aprendizaje automático de Knowledge Studio a comprender el idioma del dominio. En el aprendizaje automático, un diccionario agrupa palabras y frases que tienen algo en común. Una entrada del diccionario no significa que todas las palabras de la entrada signifiquen lo mismo, sino que un modelo tratará a las palabras de forma equivalente. Un diccionario es una lista de palabras y frases que son equivalentes para fines de extracción de información, lo que significa que son intercambiables para los fines de identificación de menciones de entidades y de relaciones. Tenga en cuenta este ejemplo: una entrada de diccionario contiene los siete días de la semana. Para anotar un documento, un anotador humano asigna el tipo de entidad DAY_OF_WEEK a menciones de *Lunes* y *Viernes* en el texto. Dado que el diccionario equipara los siete días de la semana, ayudará a garantizar que un modelo de aprendizaje automático anote correctamente apariciones de *Martes*, *Miércoles*, y el resto de los días de la semana en documentos no vistos en tiempo de ejecución. Además, equiparar estas palabras también beneficia la extracción de información en el texto circundante. Lo que aprende el modelo de aprendizaje automático de ejemplos de entrenamiento sobre los textos cerca de *Lunes* y *Viernes* se aplica a los textos que el modelo de aprendizaje automático ve cerca de otros días de la semana porque el diccionario indica que estos términos son equivalentes para fines de extracción de información.

Nota: No es necesario crear un diccionario que contenga información de los días de la semana. Se crean varios diccionarios de finalidad general como este en la aplicación. Otros diccionarios incorporados incluyen países, nombres de lugares, palabras de números, animales, plantas, enfermedades, palabras de medida (como *onza* y *metro*), y palabras de título de saludo (como *Mr.* (Señor) y *Mrs.* (Señora)). No puede inhabilitar ni editar diccionarios incorporados.

Evite añadir entradas que tengan varios significados. Por ejemplo, en un dominio sobre automovilismo, tiene sentido incluir el término *banco*, que hace referencia a una característica vial, solo si las instituciones financieras no se tratan también frecuentemente en el texto. Si se dan ambos sentidos de la palabra con frecuencia en los documentos de origen, será mejor dejarla fuera de ambos tipos de diccionarios: el diccionario que está asociado con características viales, y el que está asociado con las instituciones financieras.

Puede crear diccionarios en Knowledge Studio añadiendo manualmente entradas individuales. Knowledge Studio también da soporte a la capacidad de cargar varios tipos de archivos de diccionario.

Los diccionarios se utilizan en un par de formas, todas opcionales. Los utiliza el modelo de aprendizaje automático para proporcionar palabras o frases que son equivalentes para fines de extracción de información y durante la prea notación para arrancar el esfuerzo de anotación.

- **Uso del aprendizaje automático**

El tipo de entidad que asocie con un diccionario no se utiliza para definir reglas para el modelo de aprendizaje automático. El aprendizaje automático evalúa las menciones en los documentos de forma independiente. No presupone que una mención tenga un tipo de entidad específico solo porque la mención coincida con una entrada de un diccionario asociado con tal tipo de entidad. Pone la información en la cuenta, pero la trata como una parte de información entre otras partes de información que se recopile mediante el análisis lingüístico. De hecho, si ninguno de los términos de un diccionario se da en los documentos de datos de campo, el diccionario no lo utilizará en absoluto el modelo de aprendizaje automático.

Las interfaces también dan soporte a las siguientes características de salida comunes:

- Las etiquetas del orador reconocen distintos oradores del audio en inglés de EE. UU., inglés del Reino Unido, alemán, japonés, coreano y español. La transcripción etiqueta las contribuciones de cada orador en una conversación con varios participantes. (Funcionalidad beta.)
- La detección de palabras clave identifica frases habladas que coinciden con series de palabras clave especificadas con un nivel de confianza definido por el usuario. La detección de palabras clave resulta especialmente útil cuando las frases individuales del audio son más importantes que la transcripción completa. Por ejemplo, un sistema de soporte al cliente puede identificar palabras clave para determinar cómo direccionar las solicitudes de los usuarios.
- Los resultados provisionales devuelven hipótesis progresivas a medida que progresa la transcripción. El servicio devuelve los resultados finales cuando la transcripción está completa. Los resultados provisionales solo están disponibles con la interfaz WebSocket.
- El número máximo de alternativas proporcionan posibles transcripciones alternativas. El servicio indica los resultados finales en los que tiene la mayor confianza.
- Las alternativas a palabras solicitan palabras alternativas que suenan de forma parecida a las palabras de una transcripción.
- La confianza de palabras devuelve niveles de confianza para cada palabra de una transcripción.
- Las indicaciones de fecha y hora de palabras devuelven indicaciones de fecha y hora correspondientes al principio y al final de cada palabra de una transcripción.
- El formateo inteligente convierte fechas, horas, números, valores de moneda, números de teléfono y direcciones de internet en formatos más legibles y convencionales en las transcripciones finales. En el caso del inglés de EE.UU., también puede proporcionar frases de palabras clave para incluir

determinados signos de puntuación en las transcripciones finales. El formateo inteligente recibe soporte en audio en inglés de EE. UU., japonés y español. (Funcionalidad beta.)

- La ocultación numérica oculta, o enmascara, datos numéricos de una transcripción final. La ocultación sirve para eliminar la información personal confidencial, como números de tarjetas de crédito, de las transcripciones. La característica recibe soporte en audio en inglés de EE. UU., japonés y coreano. (Funcionalidad beta.)
- El filtrado de lenguaje obsceno censura el lenguaje obsceno de las transcripciones en inglés de EE. UU.
- Tiempo de silencio de fin de frase especifica la duración del intervalo de pausa en el que el servicio divide una transcripción en varios resultados finales en respuesta al silencio.
- Dividir transcripción al final de frase indica a los servicios que dividan una transcripción en varios resultados finales para características semánticas, como por ejemplo frases. El servicio basa su comprensión de las características semánticas en el modelo de lenguaje base que utilice con una solicitud. Los modelos de lenguaje personalizado y las gramáticas también afectan a la forma en que el servicio divide una transcripción.
- Las métricas de proceso proporcionan información de temporización detallada sobre el análisis del servicio del audio de entrada.
- Las métricas de audio proporcionan información detallada sobre las características de señal del audio de entrada.

Anexo 4

Historia de IBM

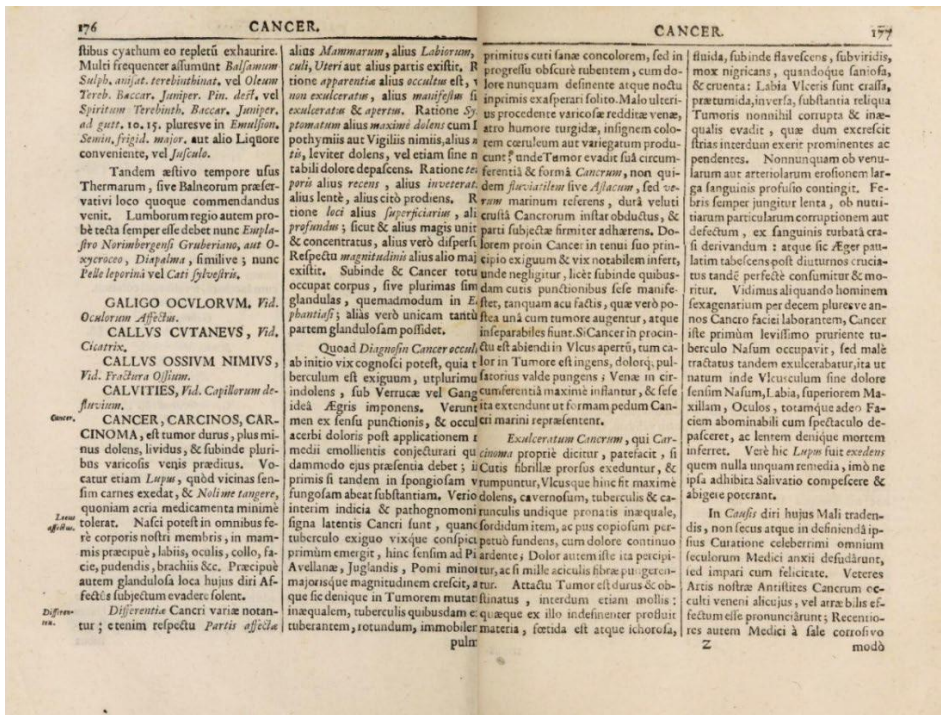
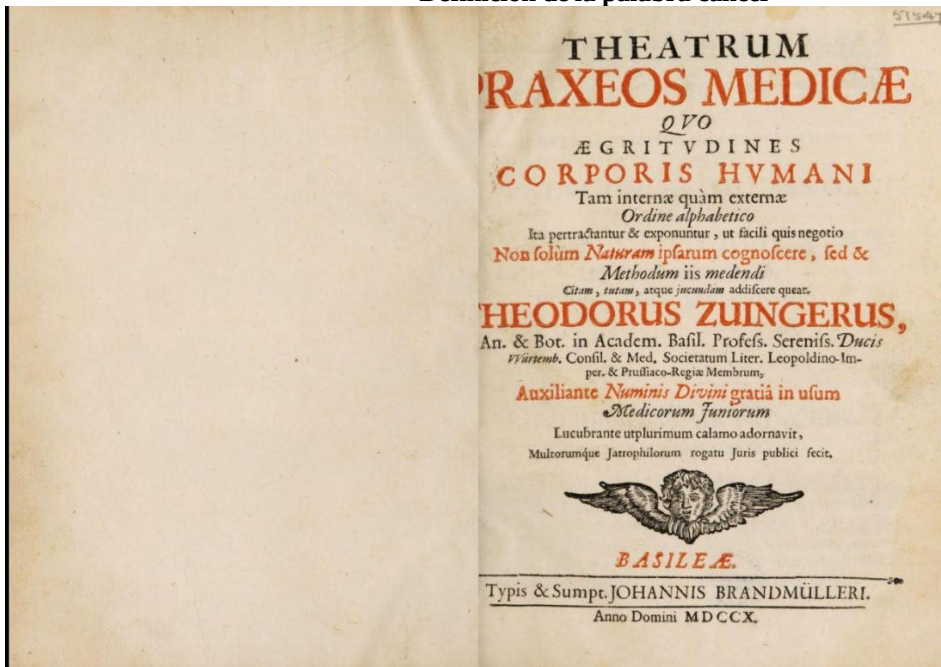
Fuente: <https://www.ibm.com/ibm/history/index.html>

Año	Acontecimiento
1889	Harlow Bundy incorpora Bundy Manufacturing Company como la primera compañía discográfica del mundo. Produce un reloj de tiempo inventado por su hermano Willard (un joyero en Auburn, Nueva York) para registrar la hora de llegada y salida de un trabajador en una cinta de papel. The Bundy Manufacturing Company se consolidará como uno de los precursores de IBM en 1902.
1890	Durante el apogeo de la Revolución Industrial, cuando Estados Unidos estaba recibiendo oleadas de nueva inmigración, la Oficina del Censo de Estados Unidos reconoció que sus métodos tradicionales de conteo serían inadecuados para medir la población en expansión. Como resultado, la Oficina patrocinó un concurso para encontrar un medio más eficiente de tabular los datos del censo. El ganador fue Herman Hollerith , hijo de un inmigrante alemán y estadístico de la Oficina del Censo, cuya máquina tabuladora de tarjetas perforadas usaba una corriente eléctrica para detectar los agujeros en las tarjetas perforadas y mantener un total acumulado de datos. Aprovechando su éxito, Hollerith formó la Compañía de Máquinas de Tabulación . El sistema de tabulación de Herman Hollerith se utiliza en el censo de EE. UU., Lo que reduce un proceso de casi 10 años a dos años y medio y ahorra \$ 5 millones.
1896	La empresa de máquinas tabuladoras de Herman Hollerith está organizada en Washington, DC, como la primera empresa de máquinas contables y tabuladoras eléctricas del mundo. La Detroit Automatic Scale Company está organizada para fabricar una báscula automática. La compañía pasará a formar parte de Computing Scale Company en 1901.
1911	El 16 de junio de 1911, Charles R. Flint (1850-1934), organiza la fusión de International Time Recording Company, Computing Scale Company y Tabulating Machine Company para formar la Computing-Tabulating-Recording Company (CTR) . La nueva empresa tiene su sede en la ciudad de Nueva York y tiene 1.300 empleados. George Fairchild se convierte en el primer presidente de la junta directiva. CTR pasará a llamarse International Business Machines Corporation en 1924.
1914	Thomas J. Watson, Sr. , se une a Computing Tabulating Recording Co. (CTR) como gerente general. Al año siguiente, Watson es elegido presidente de la empresa.
1924	Computing-Tabulating-Recording Co. se transforma en la International Business Machines Corporation (IBM) . La compañía había operado bajo el nombre de IBM en Canadá desde 1917.
1933	Se presenta el tabulador de impresión numérico IBM Type 285 , capaz de tabular 150 tarjetas por minuto. También se lanza la primera perforadora de impresión alfabética . El edificio de IBM Schoolhouse and Engineering Laboratory está dedicado en Endicott, Nueva York. La escuela presentaba a Thomas J. Watson, padre, a través del famoso “Cinco pasos hacia el conocimiento” grabado en la puerta principal: “Leer, escuchar, discutir, observar, pensar”.
1944	IBM presenta su primera computadora a gran escala, la Calculadora Automática de Secuencia Controlada (ASCC) , a la Universidad de Harvard. También conocido como Mark I, el ASCC es la primera máquina que ejecuta cálculos largos automáticamente. Con más de 50 pies de largo, dos metros y medio de alto y un peso de casi cinco toneladas, el ASCC utiliza relés electromecánicos para realizar sumas, multiplicaciones y divisiones.
1945	El Laboratorio de Computación Científica Watson se funda en la Universidad de Columbia en Nueva York, Nueva York. La renovada casa de la fraternidad en el West Side de Manhattan es el primer laboratorio de IBM dedicado a la ciencia pura. El laboratorio es el precursor de la División de Investigación de IBM, que actualmente opera instalaciones de investigación en todo el mundo.
1948	IBM presenta IBM 604 Electronic Calculating Punch . El 604 producido en serie presenta el primer ensamblaje de electrónica digital reemplazable como una unidad de la industria. Los ingenieros de servicio pueden simplemente sacar una unidad defectuosa del 604 y conectar una de repuesto, lo que reduce el tiempo de inactividad.
1952	La empresa presentó el IBM 701 , su primer ordenador grande basado en el tubo de vacío. Los tubos eran más rápidos, más pequeños y más fáciles de reemplazar que los interruptores electromecánicos del Mark I (1944). El 701 ejecutó 17.000 instrucciones por segundo y se utilizó principalmente para trabajos gubernamentales y de investigación. Pero los tubos de vacío movieron rápidamente las computadoras a aplicaciones comerciales como facturación, nómina y control de inventario. En 1959, los transistores reemplazaban a los tubos de vacío.
1955	IBM presenta la calculadora de transistores IBM 608, la primera máquina de computación totalmente de estado sólido comercializada. El 608 allanó el camino para computadoras transistorizadas más potentes y marcó la nueva dirección de diseño informático de IBM, que se formalizó en una declaración de política de 1957: “Será política de IBM utilizar circuitos de estado sólido en todos los desarrollos de máquinas. Además, no Se anunciarán nuevas máquinas o dispositivos comerciales que hagan uso principal de circuitos de tubos “.

Anexo 4 Historia de IBM

Año	Acontecimiento
1956	Thomas J. Watson, Sr. , muere a los 82 años, seis semanas después de entregar el título de director ejecutivo de IBM a su hijo, Thomas J. Watson, Jr. John McCarthy acuña el término “inteligencia artificial” en la primera conferencia de IA en Dartmouth College. (McCarthy inventaría el lenguaje Lisp). Más tarde ese año, Allen Newell, JC Shaw y Herbert Simon crearon Logic Theorist, el primer programa de software de inteligencia artificial que se ejecutaba.
1957	IBM pone a disposición de los clientes el lenguaje de programación científica FORTRAN (FORMula TRANslation). FORTRAN se convierte en el lenguaje informático más utilizado para trabajos técnicos.
1963	IBM anuncia la 7094 II , la computadora más poderosa en la línea de productos hasta la fecha; un sistema de archivo electrónico compuesto por nuevos archivos de almacenamiento en disco IBM 1302; el 1460, que procesa información casi dos veces más rápido que el 1401; y el sistema bancario 1240.
1967	Frank Rosenblatt construye el Mark 1 Perceptron, la primera computadora basada en una red neuronal que ‘aprendió’ mediante prueba y error. Solo un año después, Marvin Minsky y Seymour Papert publican un libro titulado <i>Perceptrons</i> , que se convierte tanto en el trabajo histórico sobre redes neuronales como, al menos por un tiempo, en un argumento en contra de futuros proyectos de investigación de redes neuronales.
1971	La primera aplicación operativa de reconocimiento de voz de IBM permite a los ingenieros de clientes que atienden equipos “hablar” y recibir respuestas “habladas” desde una computadora que puede reconocer alrededor de 5.000 palabras. IBM también desarrolla una terminal experimental que imprime respuestas de computadora en Braille para ciegos.
1975	IBM presenta la computadora portátil IBM 5100 , una máquina de escritorio de 50 libras que pone las capacidades de la computadora al alcance de los ingenieros, analistas, estadísticos y otros solucionadores de problemas. Más “cargable” que portátil, el 5100 puede servir como terminal para el System / 370 y cuesta entre \$ 9000 y \$ 20,000.
1976	Se instala la impresora IBM 3800 , la primera impresora que combina tecnología láser y electrofotografía. La tecnología aceleró la impresión de extractos bancarios, avisos de primas y otros documentos de gran volumen, y sigue siendo un caballo de batalla para los departamentos de facturación y cuentas por cobrar.
1980	Las redes neuronales con retropropagación (algoritmos para entrenar la red) se utilizan ampliamente en aplicaciones de inteligencia artificial.
1981	IBM anuncia la computadora personal de IBM , la más pequeña y, con un precio inicial de \$ 1,565, la computadora de IBM con el precio más bajo hasta la fecha. Un éxito inmediato, IBM PC se convierte rápidamente en el estándar de la industria, y fue una de las razones por las que la revista Time eligió la “computadora personal” como su Hombre del año de 1982. La PC también lanza toda una industria de clones, software y equipos accesorios “compatibles con IBM”.
1989	IBM se convierte en la primera empresa del mundo en ofrecer el procesador de microcomputadora más rápido de la industria, el nuevo y potente microprocesador i486 de Intel .
1992	La empresa lanza una nueva serie de ordenadores portátiles: el ThinkPad . Con una carcasa negra distintiva y un innovador dispositivo de puntero TrackPoint ubicado en el medio del teclado, el ThinkPad es un éxito inmediato y acumula más de 300 premios por diseño y calidad.
1994	La compañía forma IBM Global Network como una unidad de negocios que desarrollará y operará la red de datos y voz de alta velocidad más grande del mundo dedicada a la computación centrada en la red. IBM Global Network ya atiende a dos millones de usuarios en unas 25.000 empresas y agencias gubernamentales en más de 100 países.
1997	Deep Blue de IBM vence al entonces campeón mundial de ajedrez Garry Kasparov, en una partida de ajedrez (y revancha).
2001	IBM anuncia la disponibilidad general de ThinkPad TransNote, la primera computadora portátil de cartera del mundo que combina una computadora móvil con un bloc de notas digital y que presenta uno de los cambios de diseño más radicales de la industria desde la concha.
2007	Se crea IBM Watson, un sistema informático de propósito especial diseñado para ampliar las preguntas y respuestas profundas, análisis profundos y la comprensión de la computadora del lenguaje natural. Watson se convirtió en la primera computadora cognitiva, combinando aprendizaje automático e inteligencia artificial.
2011	IBM Watson vence a los campeones Ken Jennings y Brad Rutter en Jeopardy!
2015	La supercomputadora Minwa de Baidu utiliza un tipo especial de red neuronal profunda llamada red neuronal convolucional para identificar y categorizar imágenes con una mayor tasa de precisión que el humano promedio.
2019	IBM presentó el IBM Q System One, el primer ordenador cuántico para uso comercial . En el mismo se combina tanto la computación cuántica como “tradicional” para ofrecer un sistema de 20 qubits para su utilización en investigaciones y grandes cálculos.

Anexo 5
Definición de la palabra cáncer



Fuente: https://archive.org/details/b30550075_0001/page/n7/mode/2up [Actualizado al 15 de julio de 2022]