

UNIVERSIDAD PANAMERICANA
FACULTAD DE DERECHO
CON RVOE DE LA SEP NÚM. 985162, DE FECHA 17/08/1998

**“EL CONSENTIMIENTO INFORMADO DEL
ESTUDIANTE EN LA ERA DE LA INTELIGENCIA
ARTIFICIAL. EL MODELO DE GOBERNANZA PARA
LA CONFIANZA.”**

TESIS QUE PARA OBTENER EL GRADO EN DOCTORA EN DERECHO

PRESENTA LA

DRA. MARÍA TERESA NICOLÁS GAVILÁN

DIRECTORA DE TESIS
DRA. CLARA LUZ ÁLVAREZ GONZÁLEZ DE CASTILLA

CIUDAD DE MÉXICO, 2026

Introducción

La incorporación de la Inteligencia Artificial (IA) —y particularmente de la Inteligencia Artificial Generativa (GenAI) y, más recientemente, de la Inteligencia Artificial agéntica (IAA)— en la educación superior no constituye un fenómeno aislado ni estrictamente reciente, sino el resultado de una trayectoria de integración entre tecnologías computacionales, teorías del aprendizaje y prácticas educativas que se remonta, al menos, a la década de 1980 (Rienties et al., 2020). Sin embargo, podemos considerar que el actual estado de la IA en la educación superior requiere una atención y regulación importante por el impacto y el alcance que tiene.

El tránsito desde sistemas que generan contenido hacia agentes capaces de planificar, ejecutar acciones y decidir con grados variables de autonomía altera cualitativamente las exigencias de gobernanza, supervisión humana, protección de datos personales y consentimiento informado que pesan sobre las Instituciones de Educación Superior (IES). El reto ya no consiste únicamente en adoptar la tecnología, sino en construir sistemas institucionales capaces de demostrar que ese tratamiento se realiza con criterios verificables de transparencia, responsabilidad, supervisión humana y rendición de cuentas.

La expansión de la IA en las IES descansa sobre una infraestructura de datos que vuelve medible la vida académica y convierte cada interacción educativa en insumo de modelos que predicen, describen y recomiendan. Subcampos como el Educational Data Mining (EDM) y el Learning Analytics (LA) han consolidado el uso de datos y algoritmos para comprender y optimizar los procesos formativos (Siemens, 2013; Siemens & Gasevic, 2012). Simultáneamente, han incrementado la capacidad institucional de vigilancia y clasificación del estudiantado, con efectos potenciales sobre sus trayectorias académicas, su autonomía informacional y su privacidad (Arnold & Pistilli, 2012; Siemens, 2013). La universidad contemporánea se reconfigura así bajo una lógica de medición y gestión de datos que reordena prioridades institucionales y criterios de intervención educativa (Jones, 2019; Tsai et al., 2020). En las IES altamente digitalizadas se recopilan de manera continua datos de interacción, progreso y rendimiento a través de infraestructuras como sistemas de gestión del aprendizaje, tutoría inteligente, procesamiento de lenguaje natural y visión computacional (González-Sánchez et al., 2023; Li et al., 2020; Prinsloo, 2020; Sanabria-Navarro et al., 2023).

La promesa de personalización, predicción y mejora se sostiene, por tanto, sobre procesos sistemáticos de recolección, procesamiento y análisis de datos personales. El problema que ocupa a esta investigación es doble. Por una parte, esa promesa puede realizarse sin que medie un consentimiento verdaderamente informado. Por otra parte, la relación en que dicho consentimiento se solicita es estructuralmente asimétrica, pues quien no acepta puede perder el derecho a inscribirse en la universidad, así como gozar de ciertos accesos y herramientas digitales. De ahí que la pregunta relevante deje de ser qué puede hacer la IA, y más bien plantearse en qué reglas, garantías y límites deben incorporarse a los entornos universitarios.

Atraviesa toda la investigación una distinción que conviene enunciar desde ahora: la que media entre la licitud del tratamiento de datos y la protección efectiva de la privacidad. Un tratamiento de datos personales puede ser plenamente conforme a derecho y, sin embargo, resultar éticamente cuestionable cuando el consentimiento se agota en el cumplimiento formal de un deber de información (Mendoza Enríquez, 2020). Esta investigación muestra que las leyes actuales en México no se adaptan bien a la inclusión de la IA en la educación superior, especialmente en lo que se refiere a la protección de datos personales y al consentimiento informado de los estudiantes.

La tesis que aquí se defiende puede enunciarse en estos términos: dado que el consentimiento no puede operar como garantía suficiente en una relación estructuralmente asimétrica, y dado que el ordenamiento mexicano —a diferencia del europeo— no ha articulado los instrumentos llamados a sustituirlo, la protección efectiva de los datos personales del estudiantado debe trasladarse desde el acto de voluntad individual hacia la gobernanza institucional verificable. La confianza, entendida como articulación de carácter y competencia (Covey, 2006), constituye el criterio que permite ordenar esa gobernanza y medirla.

Coherentemente con el carácter cualitativo, jurídico y exploratorio del diseño, la investigación no formula una hipótesis susceptible de contrastación empírica, sino una proposición de trabajo, la cual cumple, dentro de este diseño, la misma función metodológica que la hipótesis de trabajo exigida por la normativa institucional: constituye el enunciado provisional que orienta la investigación y frente al cual se contrastan los hallazgos, con la diferencia de que su validación no se realiza mediante prueba estadística, sino mediante argumentación jurídica, doctrinal y comparada. Se sostiene que cuando el consentimiento informado se reduce a avisos formales, sin mecanismos operativos de comprensión, trazabilidad y control en contextos de la IA en las IES, la tutela de la privacidad se debilita. De esta proposición se deriva el criterio de suficiencia que la tesis adopta y que recorre los cuatro capítulos: un consentimiento sólo puede reputarse informado cuando es libre, específico, granular, comprensible, verificable y revocable, condiciones que no dependen del documento en que se recoge, sino de la estructura institucional que lo sostiene.

Cuatro preguntas guían el trabajo, cada una de las cuales se corresponde con uno de los objetivos específicos y con un capítulo de la tesis:

(P1) ¿Cómo incide el despliegue de sistemas de IA en la vida académica, los datos personales y la privacidad del estudiantado en las IES?

(P2) ¿Qué principios éticos pueden orientar la integración de la IA en la educación superior y sobre qué fundamento antropológico se sostienen?

(P3) ¿Cómo se configura normativamente el consentimiento informado y la protección de la privacidad ante usos de IA en educación superior en México, en comparación con la Unión Europea y Estados Unidos, y qué carencias presenta el ordenamiento mexicano?

(P4) ¿Qué dimensiones, componentes e indicadores debe integrar un modelo de gobernanza capaz de operacionalizar esas exigencias en las instituciones mexicanas, públicas y privadas, de educación superior?

El objetivo general consiste en analizar las implicaciones éticas y jurídicas del uso de la IA en la educación superior, con énfasis en la protección de datos personales y el consentimiento informado del estudiantado, y en proponer un modelo institucional que integre las dimensiones ética, normativa, operativa y educativa necesarias para una incorporación responsable y crítica de estas tecnologías en las IES mexicanas.

De este objetivo general se derivan cuatro objetivos específicos:

(O1) Caracterizar el impacto y los usos de la IA en las IES.

(O2) Examinar los principios éticos relevantes para su uso universitario y su fundamento antropológico.

(O3) Comparar las normativas nacionales e internacionales sobre IA y protección de datos, con foco en el consentimiento informado.

(O4) Proponer un modelo de gobernanza aplicable a las IES mexicanas.

La tesis se estructura en cuatro capítulos, cada uno alineado con uno de los objetivos específicos enunciados.

El Capítulo I caracteriza la incorporación de la IA en las IES, trazando su trayectoria desde las técnicas computacionales iniciales hasta los sistemas generativos y agenticos actuales, y sitúa la privacidad entre los riesgos éticos más relevantes que se derivan de la recopilación masiva de datos en contextos educativos.

El segundo capítulo examina los principios éticos que deben regir el uso de la IA en el ámbito universitario; introduce el Triángulo de la Irreducibilidad Humana como fundamento antropológico frente a la automatización, y analiza las alternativas disponibles para la evaluación ética y la gestión de riesgos, para concluir situando la confianza como criterio articulador de dicha gestión.

El Capítulo III realiza un análisis comparado del consentimiento informado y la protección de datos personales en la Unión Europea, Estados Unidos y México, a partir de una matriz de dieciocho criterios, e identifica las insuficiencias del marco normativo mexicano, así como la tensión persistente entre las exigencias formales de la ley y las garantías que esta efectivamente asegura.

El último capítulo IV presenta el Modelo de Gobernanza para la Confianza (MGC), precedido de un análisis de los instrumentos vigentes en siete instituciones mexicanas de educación superior y de los estándares ISO pertinentes, y desarrolla su implementación en las cuatro dimensiones que lo integran. Un anexo, organizado en dos tablas, contiene las fichas técnicas de los indicadores propuestos —definición, fórmula de cálculo, fuente de

datos, periodicidad, responsable, meta y riesgo asociado—, con el propósito de asegurar que el modelo resulte auditable y replicable.

La investigación se desarrolló mediante un enfoque cualitativo, documental, interdisciplinario y comparativo. El corpus de análisis integró artículos científicos, libros especializados, documentos institucionales, normas jurídicas, códigos de ética, encíclicas y documentos del magisterio, informes de organismos internacionales, recomendaciones de grupos de expertos, estándares técnicos y documentos de gobernanza vinculados con la inteligencia artificial. Se incorporaron aportes procedentes de la pedagogía, la educación, la ingeniería, la filosofía y el derecho, seleccionados conforme a criterios de pertinencia temática, autoridad académica, actualidad, procedencia institucional y relación directa con el objeto de estudio.

Las fuentes fueron examinadas mediante análisis crítico de contenido y comparación sistemática.

La literatura disponible aborda el problema de manera fragmentaria y diversas perspectivas distintas que no siempre dialogan entre sí. Se han agrupado las fuentes analizadas en cinco conjuntos.

Un primer conjunto de trabajos, de procedencia mayoritariamente anglosajona, ha documentado con solidez la insuficiencia del modelo de notificación y elección: la autogestión de la privacidad presupone una capacidad de lectura, comprensión y evaluación que la complejidad de los entornos digitales vuelve ilusoria (Solove, 2013), la protección no puede descansar en autorizaciones formales ni en una supervisión institucional meramente teórica (Zeide, 2016) y el control individual necesita combinarse con limitación de finalidad y responsabilidad institucional (Tene & Polonetsky, 2013; Reidenberg & Schaub, 2018). Esta línea, sin embargo, se detiene en el diagnóstico y no propone arquitecturas institucionales concretas. En México, las contribuciones de Mendoza Enríquez (2020), Tenorio Cueto (2019, 2026) y Quijano Decanini (2022) amplían este análisis y señalan la diferencia entre lo que dice la ley y lo que es ético en el tratamiento, aunque tampoco proponen maneras de ponerlo en práctica.

Un segundo conjunto lo integran documentos que formulan principios éticos para el uso de la IA en educación (Dignum, 2019; Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023; UNESCO, 2021). Su valor orientador es incuestionable, pero, como advierte Chan (2023), tales principios permanecen de naturaleza predictiva y prescriptiva mientras no se validen mediante implementación real, y resultan insuficientes si no se traducen en procedimientos, responsabilidades y capacidades operativas dentro de las universidades.

El análisis legal y normativo, constituye el tercer conjunto, se centró en los ordenamientos de la Unión Europea (UE), Estados Unidos de Norteamérica (EUA) y Estados Unidos de México (México). La investigación se delimitó de enero de 2010 al 31 de julio de 2026, fecha esta última de corte para la disponibilidad de las fuentes normativas, institucionales y

académicas consultadas. La elección de este período se debe a que en él se crean las bases para estudiar el consentimiento y el uso de datos personales en la educación digital, especialmente el Reglamento General de la UE para la protección de Datos (RGPD), las directivas 5/2000 y la Ley de IA (IA Act). Conviene precisar que esta última no opera como un hito único de plena entrada en vigor, sino como un régimen de aplicación escalonada. Entró en vigor el 1 de agosto de 2024. Por lo que se refiere a la legislación de EUA nos centramos en dos legislaciones la Ley de Derechos de Privacidad de California (CPRA) y la Ley de Derechos Educativos y Privacidad Familiar (FERPA). Para México se consideran los instrumentos jurídicos aplicables y sus actualizaciones vigentes durante el periodo delimitado. Esto incluye la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) y la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO), ambas de 2025. También se considera la reforma publicada en noviembre de 2025 y la Ley General de Educación en su reforma de enero de 2026. Cualquier reforma, criterio o lineamiento publicado con posterioridad queda fuera del corpus analizado y se reconoce expresamente como límite del estudio.

Un cuarto conjunto, de naturaleza técnica, se centra en sistemas de gestión y estándares de cumplimiento capaces de convertir obligaciones en procesos auditables —singularmente las normas ISO/IEC 27701, 27018, 29184, TS 27560 y 42001 (ISO, 2019a, 2019b, 2020, 2023; Lachaud, 2020). Estas normativas no fueron concebido para el ámbito educativo, pero los proveedores de sistemas de IA para IES las aplican.

El quinto conjunto son dos encuestas de percepción: la realizada por el Observatorio Interinstitucional de IA en la Educación Superior en México (OIIAES, 2026) y la encuesta de la Secretaría de Educación Pública en el 2025 denominada: Encuesta Nacional sobre usos y percepciones de la Inteligencia Artificial Generativa (ENIAG, 2025)

El último conjunto corresponde a los marcos, políticas y lineamientos emitidos por algunas IES mexicanas para orientar el uso de la IA. Se conformó una muestra intencional de siete instituciones mexicanas de educación superior que habían expedido instrumentos de gobernanza de la IA entre 2023 y 2026, seleccionadas por criterios de diversidad de régimen jurídico —públicas y privadas—, de cobertura territorial y de disponibilidad pública del instrumento.

Por lo tanto, la investigación no solo se centró en las leyes, sino que también incluyó puntos de vista éticos, educativos, técnicos y sociales para crear una interpretación completa y bien fundamentada del problema. En el caso de nuestro país encontramos que no hay una articulación de las dimensiones jurídica, organizacional, operativa, pedagógica y formativa que exige la operacionalización del consentimiento informado en entornos universitarios mediados por IES. En ese vacío se sitúa la presente investigación.

La tesis formula cinco aportaciones originales que conviene anunciar para facilitar su valoración:

1. El Triángulo de la Irreductibilidad Humana —empatía, juicio ético y pensamiento divergente— como criterio antropológico que identifica aquello que los sistemas de IA no pueden sustituir y que, por ello mismo, opera como límite a su autonomía creciente y como fundamento de la supervisión humana significativa.
2. La confianza como criterio institucional verificable, y no únicamente como valor declarativo. Según la diferencia entre carácter y competencia (Covey, 2006), la investigación convierte la confianza en pruebas, controles e indicadores que pueden ser revisados, lo que permite valorar si una institución merece la confianza que dice tener.
3. La tabla que presenta el análisis comparado las normativas nacionales, de la UE y EUA en materia de protección de datos, con foco en el consentimiento informado. El análisis se realizó mediante una matriz de dieciocho criterios agrupados en seis ejes de análisis: la arquitectura normativa, la posición del consentimiento en la estructura de licitud, el estándar de validez exigido, el tratamiento específico de la inteligencia artificial, la regulación de la cadena de proveedores y la arquitectura de garantía. La matriz permitió superar la mera yuxtaposición descriptiva de los tres regímenes y precisar, en cada eje, la carga que cada ordenamiento hace recaer sobre la figura del consentimiento y los mecanismos que arbitra para sustituirla cuando resulta insuficiente.
4. El Modelo de Gobernanza para la Confianza (MGC), que es la idea principal de esta investigación, une las áreas de gobernanza, operativa-tecnológica, pedagógica y formativa. Incluye formas de asegurarse que se cumplan las normas internacionales y se lleva a cabo a través de indicadores que se pueden verificar, como el consentimiento, la confianza y la capacidad de hacer auditorías.
5. El análisis de los marcos y lineamientos desarrollados por las IES mexicanas, constituye una aportación empírica propia de la investigación y permitió identificar tanto las convergencias de principio entre las instituciones como la ausencia generalizada de mecanismos verificables para la evaluación de riesgos, la gestión de proveedores tecnológicos, y la operacionalización del consentimiento informado. Cada documento fue clasificado conforme a diez descriptores: instrumento, año, instancia emisora, destinatario declarado, naturaleza del instrumento, jerarquía y alcance, unidad de aplicación, principios enunciados y herramientas institucionales asociadas.

Esta investigación invita a considerar la incorporación de la IA en la educación superior no como una promesa de eficiencia, sino como un desafío de gobernanza, derechos y confianza que puede resumirse en cuatro preguntas: quién decide, con qué datos, bajo qué reglas y con qué posibilidad real de comprensión y control por parte del estudiantado. La tesis presenta una idea y un enfoque para afirmar que la integración responsable de estas tecnologías

depende más de la capacidad de las instituciones para gestionarlas que de la tecnología en sí. Esto se logra a través de mecanismos claros de transparencia, supervisión humana, protección de datos, educación y confianza.

Una adopción responsable y ética de la IA en las IES se debe considerar siempre en beneficio de la persona y la sociedad, como ha señalado la Iglesia Católica “no basta con que una IA funcione: deben evaluarse fines y medios para asegurar que respetan la dignidad humana y promueven el bien común” (Dicasterio para la Doctrina de la Fe & Dicasterio para la Cultura y la Educación, 2025, p. 8).

Declaración sobre el uso de herramientas de inteligencia artificial: en la elaboración de esta tesis, la autora empleó como apoyo instrumental las siguientes herramientas de inteligencia artificial: ChatGPT (OpenAI, 2026), mediante un agente personalizado, para el diseño gráfico y para la revisión de estilo y redacción; Claude Sonnet 5 (Anthropic, 2026), para el diseño gráfico, para la revisión de estilo y ortográfica, para la verificación del cumplimiento formal y de citación conforme a la Guía de la Facultad de Derecho, y como apoyo en la búsqueda bibliográfica; Perplexity AI (2026), como apoyo en la búsqueda bibliográfica; y Gemini (Google, 2026), como apoyo en la estructuración conceptual de los marcos teóricos, la elaboración de tablas a partir de textos de la autora y la optimización estilística de los borradores. En todos los casos, tras el uso de estas herramientas, la autora revisó, analizó críticamente y editó el material resultante, asumiendo la responsabilidad plena sobre el contenido, el análisis jurídico y las conclusiones expresados en este trabajo.

Lista de acrónimos y siglas

Acrónimos técnicos (inglés–español)

AGI	<i>Artificial General Intelligence / Inteligencia Artificial general</i>
AI	<i>Artificial Intelligence / Inteligencia Artificial</i>
ANI	<i>Artificial Narrow Intelligence / Inteligencia Artificial estrecha</i>
APIS	<i>Application Programming Interface / Interfaz de Programación de Aplicaciones</i>
ASI	<i>Artificial Superintelligence / Superinteligencia artificial</i>
BD	<i>Big Data / Macrodatos (datos masivos)</i>
CV	<i>Computer Vision / Visión por computadora</i>
DA	<i>Data Analytics / Analítica de datos</i>
DL	<i>Deep Learning / Aprendizaje profundo</i>
EDM	<i>Educational Data Mining / Minería de datos en la educación</i>
GenAI	<i>Generative Artificial Intelligence / Inteligencia Artificial Generativa</i>
GPT	<i>Generative Pre-trained Transformer / Transformador generativo preentrenado</i>
HTML	<i>HyperText Markup Language / Lenguaje de Marcado de Hipertexto</i>
ITS	<i>Intelligent Tutoring Systems / Sistemas de tutoría inteligente</i>

IA	<i>Inteligencia Artificial</i>
IAED	<i>Inteligencia Artificial en la Educación</i>

IoT	<i>Internet of Things / Internet de las cosas</i>
LA	<i>Learning Analytics / Analítica de aprendizaje</i>
LMS	<i>Learning Management Systems / Sistemas de gestión del aprendizaje</i>
ML	<i>Machine Learning / Aprendizaje automático o aprendizaje de máquina</i>
NLP	<i>Natural Language Processing / Procesamiento del lenguaje natural</i>
NN	<i>Neural Networks / Redes neuronales</i>
XAI	<i>Explainable Artificial Intelligence / Inteligencia Artificial explicable</i>

Otros acrónimos y siglas

Sigla	Significado
ACE	<i>Always Center Educators/ Siempre centrados en los educadores</i>
ACNUDH	<i>Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos</i>
AI ACT	<i>Artificial Intelligence Act/ Ley de Inteligencia Artificial</i>
AI RMF	<i>AI Risk Management Framework / Marco de Gestión de Riesgos de IA</i>
ANUIES	<i>Asociación Nacional de Universidades e Instituciones de Educación Superior</i>
ARCO	<i>Derechos de Acceso, Rectificación, Cancelación y Oposición</i>
CCPA	<i>California Consumer Privacy Act / Ley de Privacidad del Consumidor de California</i>
CE-IA	<i>Comité de Ética en IA</i>
CEPD	<i>Comité Europeo de Protección de Datos</i>
CPRA	<i>California Privacy Rights Act / Ley de Derechos de Privacidad de California</i>
DOF	<i>Diario Oficial de la Federación</i>
DPA	<i>Data Protection Authority / Autoridad de protección de datos</i>
DPIA	<i>Data Protection Impact Assessment / Evaluación de Impacto en Protección de Datos</i>
DPO	<i>Data Protection Officer / delegado/a de Protección de Datos</i>
EDPB	<i>European Data Protection Board / Comité Europeo de Protección de Datos</i>
EDRi	<i>European Digital Rights/Red de Derechos Digitales Europeos</i>
EE. UU.	<i>Estados Unidos (abreviatura en español)</i>
FERPA	<i>Family Educational Rights and Privacy Act / Ley de Derechos Educativos y Privacidad Familiar</i>
GDPR	<i>General Data Protection Regulation / Reglamento General de Protección de Datos</i>
HRC	<i>Human Rights Council / Consejo de Derechos Humanos</i>
IES	<i>Instituciones de Educación Superior</i>
ISO	<i>International Organization for Standardization / Organización Internacional de Normalización</i>
KPI	<i>Key Performance Indicators / Indicadores clave de desempeño</i>
LFPDPPP	<i>Ley Federal de Protección de Datos Personales en Posesión de los Particulares</i>
LGPDPSO	<i>Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados</i>
MGC	<i>Modelo de Gobernanza para la Confianza</i>
NIST	<i>National Institute of Standards and Technology / Instituto Nacional de Estándares y Tecnología</i>
NOM	<i>Normas Oficiales Mexicanas</i>
OCDE	<i>Organización para la Cooperación y el Desarrollo Económicos (OECD)</i>
ODS	<i>Objetivos de Desarrollo Sostenible</i>
OEA	<i>Organización de los Estados Americanos</i>
ONU	<i>Organización de las Naciones Unidas</i>

PIB	<i>Producto Interno Bruto</i>
PII	<i>Personally Identifiable Information / Información personal identificable</i>
RGPD	<i>Reglamento General de Protección de Datos (UE) 2016/679</i>
RIPD	<i>Red Iberoamericana de Protección de Datos</i>
RMF	<i>Risk Management Framework / Marco de Gestión de Riesgos</i>
STEM	<i>Science, Technology, Engineering and Mathematics / Ciencia, Tecnología, Ingeniería y Matemáticas</i>
TD	<i>Transformación Digital</i>
UE	<i>Unión Europea</i>
UNESCO	<i>Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura</i>

ÍNDICE

Introducción	3
Lista de acrónimos y siglas	11
Otros acrónimos y siglas	13
CAPÍTULO 1. Impacto y Usos de la Inteligencia Artificial en Instituciones de Educación Superior	19
1. 1. ¿Cómo Hemos Llegado al Estado Actual? Las Revoluciones Tecnológicas	19
1.2. Conceptualización de la Inteligencia Artificial (IA)	21
1.3. Evolución Técnica y Nuevos Paradigmas: de lo Predictivo a lo Autónomo	23
1.3.1. Técnicas fundacionales: Machine Learning, Neural Networks y Deep Learning	27
1.3.2. Arquitectura y representación del conocimiento: de la lógica simbólica a la representación multimodal	29
1.3.3. Inteligencia Artificial Generativa (GenAI)	31
1.3.4. Inteligencia Artificial Agéntica (IA Agéntica)	32
1. 4. Campos de Acción Tecnológica	33
1. 4. 1. <i>Natural Language Processing</i> (NLP)	33
1. 4. 2. <i>Computer Vision</i> (CV)	34
1. 5. El ecosistema de datos en la Educación Superior. <i>Data Analytics</i>	36
1. 5. 1. <i>Big Data</i> y Analítica	36
1. 5. 2. <i>Learning Analytics</i> (LA) y <i>Educational Data Mining</i> (EDM)	37
1.6. Aplicaciones Prácticas de la IA en la Educación Superior	40
1. 6. 1. <i>Learning Management Systems</i> (LMS) y Tutoría Inteligente (ITS)	41
1. 6. 2. Automatización de evaluaciones	43
1. 6. 3. Gestión Educativa y Administrativa	44
1. 6. 4. Investigación Académica	46
1. 7. Riesgos y Desafíos Éticos	48
1.7.1. Acceso Equitativo	48
1.7.2. Sesgos y Discriminación	48
1.7.3. Calidad Educativa	49
1.7.4. Privacidad y Seguridad de Datos	50
CAPÍTULO 2. Principios Éticos de la IA y su Marco de Aplicación en Instituciones de Educación Superior.	51
2. 1. La IA como una Herramienta al Servicio de la Persona	51
2. 1. 1. El Triángulo de la Irreductibilidad Humana	52
2. 2. Principios de Uso Ético de la IA en la Educación	53
	15

2. 2. 1. Centrada en la Persona	54
2. 2. 2. Privacidad y Seguridad de los Datos	56
2. 2. 3. Transparencia y Explicabilidad	57
2. 2. 4. Responsabilidad y Rendición de Cuentas	60
2. 2. 5. Gobernanza y Supervisión	61
2. 2. 6. Inclusión y Equidad	63
2. 2. 7. Sostenibilidad y Proporcionalidad	65
2. 2. 8. No Maleficencia y Prevención de Daños	66
2. 2. 9. Supervisión Humana	68
2. 2. 10. Alfabetización en IA	69
2. 3. Premisa Fundamental: la Centralidad del Docente	71
2. 4. Marcos de Evaluación Ética y Gestión de Riesgos	73
2. 5. Un Modelo Ético Iterativo que Pone al Docente en el Centro	76
2. 6. Confianza como Valor Transversal	77
2. 7. La Confianza y los Principios Éticos de Uso de la IA	78
Capítulo 3. El Consentimiento Informado del Estudiante. Comparación de los Sistemas de la UE, EUA y México.	81
3. 1. Derecho a la Privacidad en la era de la IA	81
3. 2. Consentimiento Informado: Origen y Evolución	84
3. 3. El Consentimiento Informado en la ONU	86
3. 4. El Consentimiento Informado en la Legislación Europea	88
3. 4. 1. El Reglamento General de Protección de Datos (RGPD)	88
3. 4. 2. Directrices 5/2020 sobre el Consentimiento Informado	90
3. 4. 3. Ley de la IA (AI Act)	91
3. 4. 4. El Paquete Digital Omnibus 2026	95
3. 4. 5. La Doctrina Especializada	97
3. 5. El Consentimiento Informado en la Legislación Norteamericana	99
3. 5. 1. Ley de Derechos de Privacidad de California (CPRA)	100
3. 5. 2. Ley de Derechos Educativos y Privacidad Familiar (FERPA)	102
3. 5. 3. La Doctrina Especializada	103
3. 6. El Consentimiento Informado en México	105
3. 6. 1. Ley Federal de Protección de Datos Personales en Posesión de los Particulares	106
3. 6. 2. Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados	108
3. 6. 3. El Consentimiento de los Estudiantes: Comparación entre IES Públicas y Privadas	108
	16

3. 6. 4. Transferencias, Remisiones y Contratación de Proveedores de IA	112
3. 6. 5. La Reconfiguración de las Autoridades Garantes	115
3. 6. 6. La Ley General de la Educación	116
3. 6. 7. La Doctrina Especializada	117
3. 7. Análisis Comparativo de las Legislaciones de UE, EUA y México.	119
3.7.1. La Arquitectura Normativa: Integración, Fragmentación y Dualidad	119
3.7.2. La Posición del Consentimiento en la Estructura de Licitud	120
3.7.3. El Estándar de Validez del Consentimiento	121
3.7.4. El tratamiento normativo de la IA	122
3.7.5. La Cadena de Responsabilidad	123
3.7.6. La Arquitectura de Garantía	123
Capítulo 4. Propuesta del Modelo de Gobernanza para la Confianza (MGC) para la incorporación adecuada de la IA en educación superior.	129
4. 1. El Marco de Políticas de Educación Ecológica de IA	130
4. 1. 1. La Dimensión de Gobernanza	131
4. 1. 2. La Dimensión Operativa y Tecnológica	131
4. 1. 3. La Dimensión Pedagógica	132
4. 2. Integración de Marcos complementarios al Modelo Ecológico de Chan	133
4.3. Gobernanzas de las IES Mexicanas y Estándares ISO.	134
4.3.1. Gobernanza Institucional en las IES Mexicanas	134
4.3.2. Estándares ISO	136
4. 4. El Modelo de Gobernanza para la Confianza (MGC)	140
4. 5. Implementación del Modelo de Gobernanza para la Confianza (MGC)	145
4. 5. 1. Dimensión de Gobernanza	146
4. 5. 2. Dimensión Operativa	153
4. 5. 3. Dimensión Pedagógica	157
4. 6. Operacionalización del MGC	163
4. 7. KPI de Consentimiento, Confianza y Auditabilidad	166
Conclusiones	169
Bibliografía	177
ANEXO 1	205

CAPÍTULO 1. Impacto y Usos de la Inteligencia Artificial en Instituciones de Educación Superior

En cumplimiento con el primer objetivo específico de esta investigación, el presente capítulo tiene como propósito caracterizar el impacto y los diversos usos de la IA en las IES. Esta descripción es muy importante porque es la base técnica y práctica que se necesita para analizar luego las implicaciones éticas y legales que requiere el objetivo principal de la tesis. Comprender cómo se despliegan aplicaciones como la analítica del aprendizaje, los sistemas de tutoría inteligente y la gestión datificada de la vida académica permite identificar con precisión los puntos de tensión respecto a la protección de datos personales y el ejercicio del consentimiento informado en el ecosistema universitario contemporáneo (Rienties et al., 2020; Kwet & Prinsloo, 2020).

Las instituciones educativas han evolucionado con la digitalización y la datificación, integrando la IA en diversas áreas de la enseñanza y la gestión académica. Tecnologías como el *Internet of Things* (IoT), el *Big Data* (BD) y la IA han permitido transformar las aulas en entornos inteligentes, personalizar el aprendizaje y optimizar la administración educativa (Uskov et al., 2017).

Una Institución de Educación Superior (IES) que emplea IoT, IA y BD se convierte en un entorno propicio para el desarrollo de un ecosistema inteligente (*smart*), lo que da paso a conceptos como *smart learning*, *smart classroom*, *smart campus* y, en última instancia, *smart university* (Min-Allah & Alrashed, 2020; Uskov et al., 2017).

1. 1. ¿Cómo Hemos Llegado al Estado Actual? Las Revoluciones Tecnológicas

Desde la primera Revolución Industrial hasta la era digital, la tecnología ha desempeñado un papel clave en la educación. Aunque la tecnología ha estado presente en la vida del ser humano durante siglos, en las últimas décadas ha adquirido un rol determinante en el estilo de vida de las personas, las organizaciones y los Estados. Sus aplicaciones abarcan diversos ámbitos, desde la economía hasta la medicina, el entretenimiento y la gestión social. Actualmente, resulta difícil encontrar un aspecto de la vida cotidiana que no esté influenciado por los avances tecnológicos (Roser et al., 2023).

La relación entre la tecnología y el cambio social puede analizarse a través del marco de las revoluciones industriales. En los últimos tres siglos, la humanidad ha experimentado cuatro grandes revoluciones industriales, cada una de las cuales ha redefinido la producción, el trabajo y la sociedad en general (Schwab, 2017).

La Primera Revolución Industrial, durante los siglos XVIII y XIX, se caracterizó por la mecanización impulsada por la máquina de vapor y el desarrollo de industrias como la textil

y la metalúrgica. Este periodo marcó la transición de una economía agraria a un modelo industrial y urbano (Mokyr, 1990).

La Segunda Revolución Industrial, también conocida como la Revolución Tecnológica, se desarrolló entre 1870 y 1914; estuvo impulsada por el uso del acero, el petróleo y la electricidad. La aparición del motor de combustión interna y la expansión de las redes eléctricas transformaron profundamente los sistemas de producción y transporte (Freeman & Lou, 2001).

La Tercera Revolución Industrial, que comenzó en la segunda mitad del siglo XX, se conoce como la Revolución Digital porque se identificó con la transición de dispositivos mecánicos y analógicos a tecnologías digitales. La computación y las telecomunicaciones facilitaron la automatización de procesos y la globalización de la información (Pérez, 2010). La digitalización se consolidó como el eje central de esta transformación, promoviendo una interconectividad global sin precedentes (Castells, 2006; Rifkin, 2011).

La Cuarta Revolución Industrial, definida por Schwab (2017), integra tecnologías avanzadas como la IA, la robótica y el IoT, fusionando los ámbitos físico, digital y biológico. Esta revolución ha dado paso a un nuevo paradigma donde la automatización y la toma de decisiones basada en datos reconfiguran sectores clave, incluida la educación.

En los últimos años, algunos académicos han sugerido la existencia de una Quinta Revolución Industrial, caracterizada por la colaboración entre humanos y máquinas en lugar de una automatización total. A diferencia de la cuarta revolución, que prioriza la eficiencia tecnológica, esta nueva fase busca potenciar la creatividad y las capacidades humanas mediante el uso de IA y sistemas adaptativos (Porcelli & Martínez, 2020).

Este nuevo paradigma se basa en la combinación de la IA con las capacidades cognitivas humanas dentro de los entornos laborales, promoviendo una sinergia en la que las máquinas complementan las habilidades humanas en lugar de sustituirlas. Esto permite ampliar el potencial de la sociedad en ámbitos clave como la sostenibilidad, la inclusión y el desarrollo tecnológico con enfoque ético.

Independientemente de si nos encontramos en la transición entre la Cuarta y la Quinta Revolución Industrial, o incluso en el umbral de una Sexta, lo cierto es que la IA es una tecnología transformadora, transversal y de previsible alcance sistémico (Torrijos & Sánchez, 2023). Se ha consolidado como un agente transformador en múltiples ámbitos, incluido el educativo.

1.2. Conceptualización de la Inteligencia Artificial (IA)

La IA es un campo en constante evolución que aún carece de una definición única y consensuada. John McCarthy, pionero en la disciplina, la definió como la intersección entre ciencia e ingeniería con el propósito de desarrollar máquinas capaces de emular la inteligencia humana (Pedreño Muñoz *et al.*, 2024). Para Russell & Norvig (2004, 2), la IA es una ciencia que persigue crear sistemas que piensen o actúen como humanos, o que lo hagan racionalmente.

Así, la Organización para la Cooperación y el Desarrollo Económico (OCDE) define la IA como “un sistema basado en máquinas que, con objetivos explícitos o implícitos, infiere a partir de la información que recibe (*input*) cómo generar resultados tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales” (OCDE, 2024, pp. 3-4).

Algunos autores destacan el concepto de inteligencia en las máquinas. En este sentido, González-Campos *et al.* (2024) definen la IA como la capacidad de las máquinas para ejecutar tareas que requieren inteligencia humana, como el aprendizaje, la resolución de problemas y la toma de decisiones. En ese mismo sentido Popenici, considera que la IA son “sistemas informáticos capaces de realizar procesos similares a los humanos, como aprender, adaptarse, sintetizar, autocorregirse y utilizar datos para tareas de procesamiento complejas” (citado por Crompton & Burke, 2023 p. 2).

Sin embargo, desde una perspectiva más humanista, documentos como *Antiqua et Nova* advierten que, a pesar de su potencia en el procesamiento de datos, la IA sigue siendo funcional. Carece de una comprensión integral de la realidad, así como de discernimiento moral y de la capacidad de establecer relaciones interpersonales. Estos son atributos propios de la inteligencia humana (Dicasterio para la Doctrina de la Fe, 2025).

En el contexto de esta investigación, se propone la siguiente definición de IA:

“Conjunto de sistemas tecnológicos diseñados para procesar información, reconocer patrones y generar respuestas funcionales a problemas específicos, imitando ciertas capacidades cognitivas humanas, con la finalidad de asistir en la toma de decisiones y optimización de tareas, siempre bajo el respeto a la dignidad humana y con aplicación de principios éticos”.

Las diferentes definiciones de IA permiten identificar dos enfoques principales para clasificarla. El primero se basa en la emulación de la inteligencia humana a través de procesos cognitivos, mientras que el segundo se enfoca en la automatización de tareas, lo que da lugar a una inteligencia de carácter tecnológico.

Si la IA se entiende como máquinas que emulan el comportamiento y las funciones humanas, según Arend Hintze (2018) los sistemas de IA pueden dividirse en cuatro categorías:

1. Máquinas reactivas: sistemas sin memoria que toman decisiones basadas únicamente en el contexto inmediato. Un ejemplo es *Deep Blue*, el programa de ajedrez de IBM que derrotó a Garry Kasparov en 1997 (Smith, 2006).
2. Máquinas con memoria limitada: sistemas que aprenden de experiencias pasadas para mejorar su desempeño. Ejemplos de esta categoría incluyen el reconocimiento facial y los vehículos autónomos (Hintze, 2018).
3. Teoría de la mente: aunque todavía en desarrollo, esta categoría busca que las máquinas comprendan y respondan a las emociones humanas. Su aplicación más cercana se encuentra en la IA emocional, que intenta interpretar expresiones faciales (Hintze, 2018).
4. Autoconciencia: representa el nivel más avanzado de IA, aún teórico, en el que las máquinas podrían ser conscientes de sí mismas y de su entorno. Su desarrollo implicaría significativos desafíos éticos y filosóficos (Hintze, 2018).

Desde una perspectiva tecnológica e informática, la IA se clasifica en tres niveles (Sahota & Ashley, 2019):

1. IA Estrecha (ANI, *Artificial Narrow Intelligence*): Diseñada para realizar tareas específicas dentro de un dominio concreto, como el diagnóstico médico asistido por IA.
2. IA de Propósito General (*General-Purpose Artificial Intelligence* o *GPAI*): Corresponde a modelos de IA entrenados para realizar una amplia variedad de tareas sin haber sido desarrollados para una aplicación específica. Estos modelos, también denominados modelos fundacionales (foundation models), constituyen la base de herramientas como Chat GTP, Gemini o Claude y pueden adaptarse a múltiples contextos mediante instrucciones del usuario o ajustes posteriores (Bommasani et al., 2021; UNESCO, 2023).
3. IA General (AGI, *Artificial General Intelligence*): Hace referencia a una IA con capacidades cognitivas similares a las humanas, capaz de aprender y aplicar conocimientos en diferentes contextos.
4. SuperIA (ASI, *Artificial Superintelligence*): Hipotética forma de IA que superaría la inteligencia humana en todos los aspectos, vinculada al concepto de singularidad tecnológica (Lemelshtich, 2018).

Cabe destacar que la AGI y la ASI aún no se han desarrollado y siguen siendo objeto de debate teórico.

1.3. Evolución Técnica y Nuevos Paradigmas: de lo Predictivo a lo Autónomo

La historia de la IA ha pasado por diversos ciclos de entusiasmo y estancamiento. En 1943, Warren McCulloch y Walter Pitts propusieron el primer modelo de redes neuronales artificiales, sentando las bases del aprendizaje automático. Posteriormente, en 1950, Alan Turing introdujo su célebre prueba de *Turing*, diseñada para evaluar la capacidad de una máquina para imitar la inteligencia humana (Russell *et al.*, 2004; Crompton & Burke, 2023).

En 1955, Allen Newell y Herbert Simon crearon *Logic Theorist*, considerado el primer programa de IA. Sin embargo, el consenso tácito sitúa el origen del término IA en la Conferencia de *Dartmouth*, Estados Unidos, en 1956. Aquel *Dartmouth Summer Research Project on Artificial Intelligence* concebido como una larga lluvia de ideas, reunió, entre otros, a expertos como John McCarthy, Marvin Minsky y Claude Shannon.

Las primeras demostraciones de la IA sobre su capacidad para resolver problemas matemáticos mostraron su potencial en la educación, inspirando el desarrollo de futuros sistemas de enseñanza asistida por computadora (Sanabria-Navarro *et al.*, 2023).

Durante las décadas de 1960 y 1970, la IA avanzó gracias a la mejora en la velocidad de procesamiento de las computadoras. En este período surgieron los primeros intentos de aplicar IA en la educación, como el sistema PLATO (Programmed Logic for Automatic Teaching Operations), desarrollado en la Universidad de Illinois. Este sistema permitió a los estudiantes interactuar con ejercicios personalizados y recibir retroalimentación inmediata, sentando las bases de los actuales Learning Management Systems (LMS) (Prinsloo, 2020).

Otro hito de esta época fue Eliza, un *chatbot* diseñado por Joseph Weizenbaum para simular la interacción con un terapeuta. Aunque limitada en su capacidad, Eliza inspiró el desarrollo de sistemas educativos interactivos basados en lenguaje natural. También surgieron SCHOLAR, un tutor inteligente que ajustaba la enseñanza de geografía según el conocimiento del estudiante (Tomalá De La Cruz *et al.*, 2023), y LOGO, creado por Seymour Papert para enseñar programación básica a niños, promoviendo el pensamiento computacional (Sanabria-Navarro *et al.*, 2023).

Durante el periodo de 1974 a 1980, la IA vivió un periodo de desinterés y recorte de financiamiento debido a las limitaciones tecnológicas. Es lo que se conoce como el primer invierno (Torrijos & Gutiérrez, 2023).

Del optimismo inicial se pasa a un pesimismo medido. Disminuye la inversión en proyectos, pero también el apoyo a la idea de lograr cerebros o mentes artificiales.

En la década de 1980 comenzaba lo que se suele catalogar como el segundo verano de la IA, que abarca de 1981 a 1987 (Torrijos & Giménez, 2023). Japón impulsó la IA con su Proyecto de Quinta Generación de Computadoras, lo que motivó a otras naciones a invertir en esta tecnología (Zarkadakis, 2015). En esta década, la IA dejó de enfocarse en la creación de una

inteligencia general y se orientó hacia sistemas expertos, diseñados para resolver problemas específicos en áreas como la educación. Fue en esta época cuando Geoffrey Hinton y John Hopfield hicieron avances clave en redes neuronales y aprendizaje profundo, optimizando la capacidad de la IA para procesar vastos volúmenes de datos y extraer patrones de aprendizaje (Ponce et al. 2025). Se vivió un nuevo momento de esplendor de la IA a través de la popularización del uso de los llamados sistemas expertos (Torrijos & Gutiérrez, 2023). En la educación técnica y profesional, los *Intelligent Tutoring Systems* (ITS) comenzaron a consolidarse; SOPHIE, por ejemplo, permitía a los estudiantes resolver problemas eléctricos mediante simulaciones, mientras que Mac Project ayudaba a organizar proyectos académicos y desarrollar habilidades de planificación (Vélez Rivera et al., 2024).

A finales de la década de 1990 y principios de los 2000, la IA adoptó un enfoque más empírico, utilizando la creciente disponibilidad de datos para mejorar sus modelos de aprendizaje (Russell et al., 2004). El surgimiento del BD permitió analizar grandes volúmenes de información, lo que impulsó el desarrollo de sistemas de evaluación adaptativa. Un ejemplo clave es ALEKS (*Assessment and Learning in Knowledge Spaces*), que identifica debilidades en los estudiantes y genera ejercicios personalizados, permitiendo un progreso un 30% más rápido en comparación con métodos tradicionales (Tomalá De La Cruz et al., 2023). Sin embargo, hasta bien entrados los 2000, las inversiones en IA volvieron a menguar. Llegaba el segundo invierno.

Durante décadas, la IA operó bajo el paradigma del aprendizaje supervisado y no supervisado enfocado a la clasificación, el reconocimiento de patrones y la predicción (Schmidhuber, 2015; Russell y Norvig, 2020). Estos sistemas requerían datos de entrenamiento masivos y estructurados para determinar, por ejemplo, si un correo era *spam* o no, o si un estudiante presentaba riesgo de deserción escolar basado en variables fijas (Coursera, 2025). Estos modelos eran fundamentalmente pasivos y reactivos: procesaban una entrada fija y devolvían una etiqueta o estimación de probabilidad sin capacidad de generar nuevas estructuras cognitivas o llevar a cabo acciones autónomas en el entorno virtual (Databricks, s.f.t).

El auge del internet en el siglo XXI impulsó la creación de LMS como Moodle y Blackboard, los cuales incorporaron herramientas de *Learning Analytics* (LA) para personalizar rutas de aprendizaje y predecir el rendimiento estudiantil (Rojas & Chiappe, 2024). La IA comenzó a usarse para mantener la integridad académica con herramientas como Turnitin, que combinan detección de plagio y retroalimentación automatizada.

A partir de 2010, la IA alcanzó una nueva fase de expansión gracias al Deep Learning (DL) y al *Data Analytics* (DA). Universidades como Purdue implementaron sistemas como Course Signals, que utilizaban análisis predictivo para detectar estudiantes en riesgo de deserción. Empresas como IBM desarrollaron soluciones como Watson Education, orientadas a personalizar los procesos de enseñanza-aprendizaje (Cornejo-Plaza & Cippitani, 2023). Algunos autores denominan a esta etapa la década del "tecnoentusiasmo" (Torrijos & Gutiérrez, 2023), caracterizada por el acelerado desarrollo de algoritmos, la disponibilidad

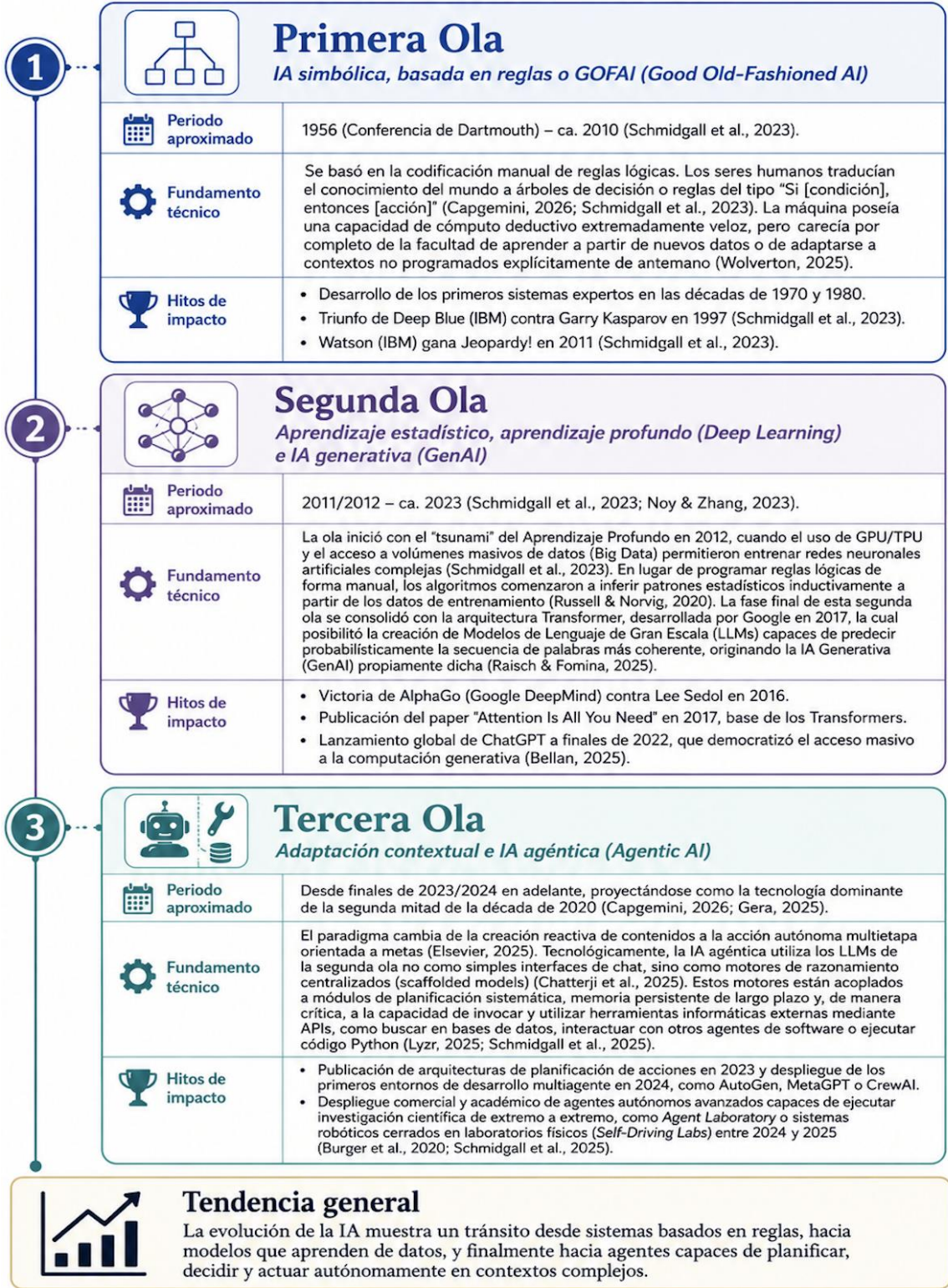
de grandes volúmenes de datos y el incremento en la capacidad de procesamiento computacional.

Durante esta etapa se produjeron avances tecnológicos que modificaron profundamente la evolución de la IA. El avance del aprendizaje profundo y la creación de nuevas estructuras para procesar el lenguaje han llevado a modelos más complejos y flexibles. Estos modelos pueden entender, generar información y realizar tareas que antes necesitaban ayuda humana (Goodfellow et al., 2016; Vaswani et al., 2017).

Como resultado de esta evolución surgieron nuevas generaciones de IA, entre ellas la IA Generativa y, más recientemente, la IA Agéntica. Estas tecnologías representan una nueva etapa en el desarrollo de la IA y han ampliado considerablemente sus posibilidades de aplicación en ámbitos como la educación, la investigación y la gestión institucional. Debido a la relevancia que tienen para esta investigación, ambas se describen con mayor detalle en los apartados siguientes (Bommasani et al., 2021; Feuerriegel et al., 2024; Bandi et al., 2025).

Con el propósito de sintetizar la evolución descrita en este apartado, se presenta la Figura 1, la cual resume las tres grandes olas de desarrollo de la IA. En ella se muestran, de manera esquemática, el período aproximado de cada etapa, sus principales fundamentos técnicos y algunos de los hitos que marcaron su evolución, desde la IA simbólica basada en reglas, pasando por el aprendizaje profundo y la IA Generativa, hasta la reciente IA Agéntica. Esta representación facilita comprender la transición tecnológica que ha dado origen a las capacidades actuales de la IA y contextualiza los conceptos y temas que se exponen en los apartados siguientes.

Figura 1. Evolución de la IA. Técnicas e hitos de impacto



Nota: Infografía generada mediante un proceso iterativo iniciado con el prompt "A partir de este texto, genera una infografía que refleje la evolución de la IA, señalando en cada ola el periodo en que se dio, el fundamento técnico y los hitos de impacto, y pon un texto de resumen al final," y refinada mediante instrucciones sucesivas de la autora sobre diseño y contenido, por ChatGPT (OpenAI, 2026) (<https://chatgpt.com/>), a partir del contenido desarrollado por la autora.

1.3.1. Técnicas fundacionales: Machine Learning, Neural Networks y Deep Learning

La IA ha transformado el ámbito educativo a través de diversas técnicas que optimizan la enseñanza y la gestión académica. Entre ellas, el *Machine Learning* (ML), las *Neural Networks* (NN) y el *Deep Learning* (DL) (Schmidhuber, 2015) han sido fundamentales para analizar grandes volúmenes de datos y generar modelos predictivos que personalizan la educación y mejoran la administración educativa (Romero y Ventura, 2020).

El *Machine Learning* (ML) es una rama de la IA que permite a los sistemas aprender y mejorar su desempeño con la experiencia, sin necesidad de programación explícita para cada tarea (Samuel, 1959; Mitchell, 1997). Se basa en la identificación de patrones en grandes volúmenes de datos, lo que permite realizar predicciones y tomar decisiones en entornos inciertos (Murphy, 2012).

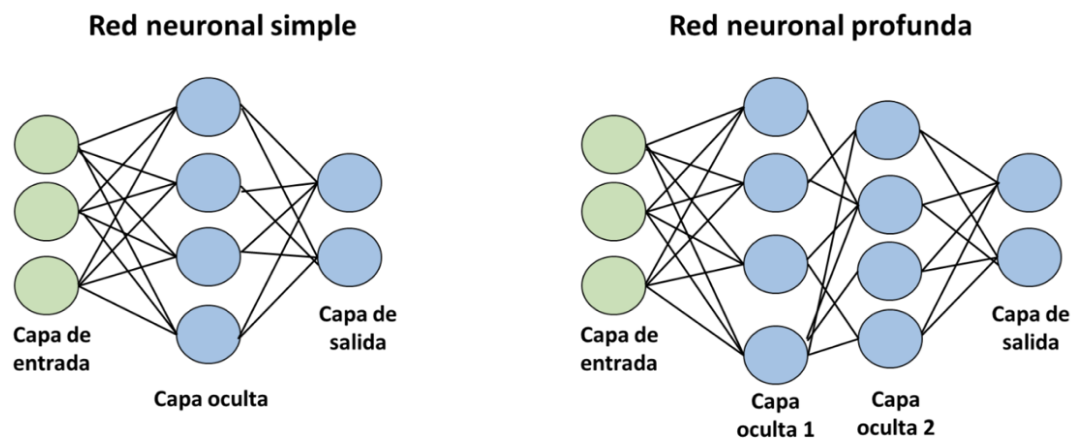
Las principales técnicas de ML incluyen el aprendizaje supervisado: Entrenamiento de modelos con datos etiquetados para hacer predicciones; el aprendizaje no supervisado: Identificación de patrones en datos sin etiquetar y; el aprendizaje por refuerzo: Optimización de decisiones mediante un sistema de recompensas y castigos (Murphy, 2012).

Las *Neural Networks* (NN) o redes neuronales, están inspiradas en la estructura del cerebro humano y se componen de neuronas artificiales organizadas en capas. Cada neurona procesa información y la transmite a la siguiente capa, permitiendo modelar relaciones no lineales en los datos y mejorar el análisis del aprendizaje estudiantil.

El desarrollo clave en este campo se produjo con el algoritmo de *backpropagation*, diseñado por Rumelhart *et al.* (1986), que optimizó el ajuste de los parámetros de la red neuronal y redujo los errores de predicción. Esta técnica ha sido aplicada en múltiples ámbitos educativos, como el reconocimiento de voz y texto en asistentes digitales, la clasificación del desempeño estudiantil y la detección de patrones de aprendizaje (Alpaydin, 2016).

Para comprender lo que es una red neuronal simple y profunda se presenta la siguiente figura:

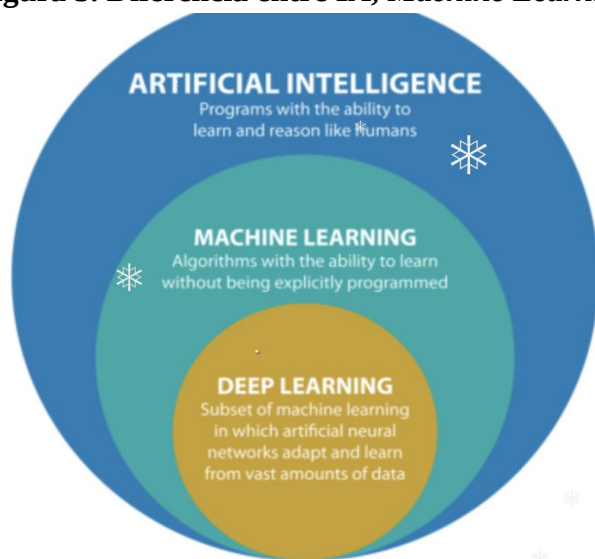
Figura 2. Redes neuronales: simple y profunda.



Nota. Tomada de *¿Qué es una red neuronal en Deep Learning?*, por KeepCoding, 2024 (<https://keepcoding.io/blog/red-neuronal-en-deep-learning/>).

El *Deep Learning* (DL) es una evolución dentro del ML que emplea DNN profundas (*deep neural networks*), es decir, redes con múltiples capas ocultas que descomponen problemas complejos en representaciones progresivamente más abstractas (Schmidhuber, 2015; Murphy, 2012). Su capacidad para analizar grandes volúmenes de datos ha permitido su integración en la educación (Lemelshtich, 2018), con aplicaciones que pueden identificar patrones de comportamiento del estudiante y generar recomendaciones personalizadas para optimizar los procesos de enseñanza (Ponce et al., 2025) e incluso adaptan el contenido educativo según el nivel de cada estudiante, proporcionando una experiencia de aprendizaje más eficiente (Uskov et al., 2017). Para tener claridad sobre las relaciones y diferencias se presenta la siguiente imagen.

Figura 3. Diferencia entre IA, Machine Learning (ML) y Deep Learning (DL).



Nota: tomada de Cinquegrana, P., y Aggarwal, S. (2018)

Estas técnicas han permitido el desarrollo de herramientas avanzadas en la educación, como plataformas de evaluación automatizada, asistentes de escritura impulsados por IA y modelos predictivos de éxito académico. Sin embargo, su implementación debe realizarse bajo principios éticos para garantizar la equidad y la transparencia en su uso (Schmidhuber, 2015; Uskov *et al.*, 2017).

El desarrollo del ML, las redes neuronales y, especialmente, del DL ha marcado un punto de inflexión en la evolución de la IA. Gracias a estas técnicas, los sistemas actuales no solo pueden reconocer patrones complejos y realizar predicciones con altos niveles de precisión, sino también procesar lenguaje natural, generar contenidos originales y ejecutar tareas cada vez más sofisticadas.

Con la llegada de la arquitectura *Transformer* en 2017 y el escalamiento de los Modelos de Lenguaje de Gran Escala (LLM), se consolidó la IA Generativa (Databricks, 2025). Su lógica operativa no es clasificar, sino crear muestras de datos totalmente novedosas a partir de la probabilidad condicional de distribución de palabras u otros elementos (Kingma & Welling, 2022).

En los últimos dos años se está desarrollando la IA Agéntica, caracterizada por su capacidad para planificar acciones, utilizar herramientas y actuar de manera autónoma orientadas al cumplimiento de objetivos específicos. Estas nuevas modalidades representan la evolución más reciente de la IA y están transformando de manera significativa los entornos educativos, administrativos y de investigación.

1.3.2. Arquitectura y representación del conocimiento: de la lógica simbólica a la representación multimodal

La representación del conocimiento constituye uno de los elementos centrales de la IA, ya que determina la forma en que un sistema almacena, organiza y utiliza la información para resolver problemas y tomar decisiones. Durante las primeras décadas de desarrollo de la IA predominó el paradigma de la lógica simbólica, basado en reglas explícitas, inferencias deductivas y representaciones formales del conocimiento. Estos sistemas ofrecían un razonamiento explícito e interpretable, pero presentaban limitaciones para enfrentar problemas caracterizados por la incertidumbre, la ambigüedad del lenguaje natural y la enorme variabilidad de los contextos reales (Russell & Norvig, 2021).

El desarrollo del aprendizaje profundo transformó este paradigma al sustituir gran parte de las reglas programadas manualmente por representaciones aprendidas automáticamente a partir de grandes volúmenes de datos. En este contexto surgieron las representaciones vectoriales (*embeddings*), mediante las cuales palabras, imágenes, documentos y otros tipos de información son codificados como vectores matemáticos dentro de espacios multidimensionales. Esta forma de representación permite que los modelos identifiquen relaciones semánticas complejas entre conceptos y posteriormente, sirvió como fundamento

para el desarrollo de los actuales modelos de lenguaje de gran escala (LLM) (Goodfellow et al., 2016; Vaswani et al., 2017).

La evolución de los modelos fundacionales permitió crear sistemas más versátiles, que pueden adaptarse a varias tareas e integrarse con herramientas externas, memorias y métodos de planificación. Estas habilidades han aumentado mucho la capacidad de la IA y han ayudado a crear sistemas más flexibles, que pueden interactuar con diferentes entornos y ayudar en procesos complicados en varios campos, como la educación (Bommasani et al., 2021; Feuerriegel et al., 2024).

En el ámbito educativo, estos avances resultan especialmente relevantes para el desarrollo de Sistemas de Tutoría Inteligente (ITS), plataformas de aprendizaje adaptativo y asistentes académicos basados en IA. La incorporación de estructuras explícitas de conocimiento permite mejorar la coherencia de las recomendaciones pedagógicas, fortalecer la transparencia de las decisiones automatizadas y reducir algunos de los riesgos asociados a la generación de información incorrecta o sesgada por parte de modelos basados exclusivamente en aprendizaje estadístico (Kasneci et al., 2023; UNESCO, 2023).

Una evolución reciente en la forma de representar el conocimiento es el uso de grafos de conocimiento; es decir, estructuras que organizan entidades y sus relaciones de forma clara para mostrar información sobre un tema específico ante relaciones semánticas, y que facilitan la recuperación de información contextualizada, mejorando la precisión y la trazabilidad de los sistemas basados en IA (Ji et al., 2022). Los grafos de conocimiento, combinados con modelos básicos, ayudan a mejorar el razonamiento de los sistemas inteligentes al permitir el acceso a fuentes de información organizadas y confiables. Esto beneficia aplicaciones educativas y científicas que necesitan más confianza y claridad (Bommasani et al., 2021; Feuerriegel et al., 2024).

Los avances en IA han impulsado el desarrollo de modelos capaces de integrar simultáneamente múltiples modalidades de información, como texto, imágenes, audio, video y otras señales digitales. Esta representación multimodal mejora la habilidad de los sistemas para entender el contexto y hacer más ricas las interacciones con los usuarios, al juntar diferentes fuentes de información en un mismo proceso de análisis. En el ámbito educativo, estas capacidades favorecen el desarrollo de entornos de aprendizaje más adaptativos y una comprensión más amplia de los procesos de enseñanza y aprendizaje. Un ejemplo de estas aplicaciones puede observarse en entornos de investigación automatizada, donde sistemas inteligentes integran múltiples fuentes de información para apoyar procesos experimentales complejos (Burger et al., 2020).

En educación superior, este enfoque ha dado lugar a la Analítica del Aprendizaje Multimodal (*Multimodal Learning Analytics*, MMLA), la cual integra diferentes fuentes de información para comprender con mayor profundidad los procesos de aprendizaje. La combinación de datos provenientes de diversas modalidades permite identificar patrones relacionados con la

participación, el compromiso y el progreso del estudiantado, favoreciendo intervenciones pedagógicas más oportunas y personalizadas (Tancredi et al., 2022). Estas capacidades también brindan nuevas oportunidades para mejorar la accesibilidad y la inclusión en la educación, pero también traen desafíos importantes sobre la privacidad, la transparencia y el uso ético de los datos (UNESCO, 2023). Estos temas se tratarán en los siguientes capítulos .

1.3.3. Inteligencia Artificial Generativa (GenAI)

La IA Generativa (Generative Artificial Intelligence o GenAI) constituye una de las evoluciones más significativas dentro del campo de la IA contemporánea. A diferencia de los sistemas tradicionales de IA, diseñados principalmente para clasificar, predecir o recomendar información a partir de datos existentes, la GenAI posee la capacidad de generar contenido nuevo y original en forma de texto, imágenes, audio, video, código informático y otros formatos digitales (Feuerriegel et al., 2024).

Su desarrollo ha sido posible gracias a los avances en Deep Learning, al incremento exponencial de la capacidad computacional y a la disponibilidad de grandes volúmenes de datos. Estos sistemas son entrenados sobre enormes conjuntos de información que les permiten identificar patrones complejos y generar respuestas coherentes a partir de los datos utilizados durante su entrenamiento (Goodfellow et al., 2016).

Uno de los avances más importantes en esta evolución fue la llegada de la arquitectura Transformer, presentada por Vaswani et al. (2017). Esta arquitectura mejoró mucho la habilidad de los modelos para manejar secuencias complejas de lenguaje natural gracias a los mecanismos de atención. Este desarrollo sentó las bases para la construcción de modelos fundacionales de gran escala (*foundation models*) y de los denominados *Large Language Models* (LLM), que actualmente constituyen la base tecnológica de gran parte de las aplicaciones de GenAI (Bommasani et al., 2021; Brown et al., 2020).

La GenAI se destaca por aprender representaciones complicadas de los datos y usar ese conocimiento para crear nuevos contenidos que mantienen una coherencia estadística con los patrones aprendidos durante el entrenamiento, sin limitarse a copiar información ya existente (Feuerriegel et al., 2024). Esta tecnología representa un cambio de paradigma debido a que amplía las capacidades humanas de producción, análisis y comunicación del conocimiento, permitiendo interacciones más naturales entre las personas y los sistemas digitales.

En el ámbito educativo, la GenAI ha adquirido una relevancia especial debido a su potencial para apoyar tanto a estudiantes como a docentes. Entre sus aplicaciones más frecuentes se encuentran la generación de materiales didácticos, la elaboración de resúmenes, la traducción de contenidos, la creación de actividades de aprendizaje personalizadas, la retroalimentación automatizada de textos académicos y el apoyo tutorial mediante asistentes conversacionales inteligentes (Kasneci et al., 2023).

Diversas herramientas de IA basadas en modelos fundacionales, como Chat GPT, Gemini, Claude y Copilot, permiten a los usuarios interactuar mediante lenguaje natural para resolver dudas, producir contenido o recibir orientación personalizada en tiempo real. Las herramientas antes mencionadas y otras plataformas basadas en modelos fundacionales permiten a los usuarios interactuar mediante lenguaje natural para resolver dudas, producir contenido o recibir orientación personalizada en tiempo real. Estas capacidades han impulsado nuevas formas de aprendizaje autónomo y de apoyo al proceso de enseñanza-aprendizaje, especialmente en contextos de educación superior y educación en línea (Chan, 2023).

La GenAI sigue siendo esencialmente reactiva: genera contenido en respuesta a instrucciones del usuario, sin perseguir objetivos por iniciativa propia ni interactuar con entornos externos salvo indicación expresa. Frente a ello, la IA Agéntica inicia tareas, prioriza metas dinámicamente, monitorea su progreso y ajusta su comportamiento mediante bucles de retroalimentación para resolver metas multi-etapa (Abou Ali et al., 2025; Deloitte, 2024).

1.3.4. Inteligencia Artificial Agéntica (IA Agéntica)

La IA Agéntica (*Agentic Artificial Intelligence* o *Agentic AI*) representa una evolución reciente de la IA basada en modelos fundacionales y sistemas generativos avanzados. La IA Agéntica representa un salto ontológico de la "generación a demanda" a la "acción autónoma orientada a objetivos" (Abou Ali et al., 2025; Elsevier, 2025). A diferencia de la GenAI, cuyo propósito principal consiste en generar contenido en respuesta a instrucciones proporcionadas por los usuarios, la IA Agéntica incorpora capacidades de razonamiento orientado a objetivos, planificación, toma de decisiones y ejecución de acciones dentro de determinados entornos operativos (Abou Ali et al., 2025; Bandi et al., 2025).

El concepto de IA Agéntica tiene sus fundamentos en la noción de agente inteligente, entendido como una entidad capaz de percibir información de su entorno, procesarla y actuar de manera orientada al cumplimiento de objetivos específicos (Russell & Norvig, 2021). Sin embargo, los avances recientes en modelos fundacionales y arquitecturas de IA, han ampliado significativamente estas capacidades, permitiendo que los agentes no solo respondan consultas, sino que desarrollen secuencias de acciones, utilicen herramientas digitales, gestionen información proveniente de diversas fuentes y adapten su comportamiento en función de los resultados obtenidos (Bommasani et al., 2021). En términos prácticos, la GenAI produce información, mientras que la IA agéntica puede producir información y actuar sobre ella mediante la interacción con herramientas, aplicaciones o sistemas externos (Bandi et al., 2025).

Los sistemas de IA Agéntica suelen integrar modelos de lenguaje de gran escala (LLM), acoplados a un módulo de planificación (*planning*), memoria a corto y largo plazo (*memory*) y un conjunto de herramientas externas (*tools* como APIs y bases de datos) (Schmidgall et al., 2025). El agente opera de forma cíclica y proactiva. Esta combinación les permite realizar

procesos complicados para alcanzar metas ya definidas, manteniendo un cierto nivel de autonomía operativa dentro de los parámetros establecidos por los usuarios o las organizaciones que los implementan (Abou Ali et al., 2025).

En el ámbito educativo, la IA Agéntica presenta posibilidades emergentes para apoyar procesos académicos, administrativos y de acompañamiento estudiantil. En lugar de responder a un prompt estático, la IA Agéntica recibe un objetivo abstracto (por ejemplo: *"monitorea la tasa de aprobación escolar y reduce de manera proactiva la tasa de abandono en la facultad de derecho"*), descompone ese fin en subtareas y actúa de manera independiente interactuando con bases de datos externas sin mediación humana constante (Abou Ali et al., 2025; IBM, 2025).

Estos sistemas podrían colaborar con docentes e investigadores en tareas relacionadas con la búsqueda de información, la organización documental, la gestión de proyectos académicos y el análisis de grandes volúmenes de datos educativos (Abou Ali et al., 2025; Bandi et al., 2025).

Para comprender mejor la GenAI y la IA Agéntica, es necesario revisar los principales campos tecnológicos que sustentan estos sistemas.

1. 4. Campos de Acción Tecnológica

En el contexto de la IA, tres campos o áreas de estudio y desarrollo han demostrado un impacto significativo en la optimización del aprendizaje y la gestión académica: *Natural Language Processing* (NLP), *Computer Vision* (CV) y *Data Analytics* (DA).

1. 4. 1. Natural Language Processing (NLP)

El NLP es una subdisciplina de la IA que permite a las máquinas comprender, interpretar y generar lenguaje humano de manera sofisticada (Jurafsky & Martin, 2020). Su objetivo principal es mejorar la interacción entre humanos y sistemas automatizados, facilitando tareas como la traducción automática, la generación de texto y la comprensión del lenguaje natural (Angeli et al., 2015).

El NLP opera en dos dimensiones clave: comprensión y generación de texto. La primera implica que los modelos de IA analicen el contexto, ambigüedades y matices del lenguaje humano mediante técnicas de análisis sintáctico y semántico (Mikolov et al., 2013). La segunda, basada en el uso de redes neuronales profundas, permite la producción de textos coherentes en respuesta a entradas específicas (Vaswani et al., 2017). Este enfoque sentó las bases para modelos más avanzados como GPT y Chat GPT, en los que la pragmática del lenguaje fue formalizada matemáticamente, dando lugar a una generación de IA capaz de comprender y generar texto con una fluidez sin precedentes (Brown et al., 2020).

La integración del NLP con ML y DL permite a los sistemas mejorar su precisión y adaptabilidad conforme procesan grandes volúmenes de datos textuales (Ruder, 2018). Aplicaciones como asistentes virtuales (Alexa, Siri, Google Assistant), sistemas de traducción automática (Google Translate) y análisis de sentimientos han transformado industrias mediante la optimización de la interacción humano-máquina. Además, plataformas como Netflix emplean NLP para personalizar recomendaciones a partir del análisis de reseñas y comentarios.

Los modelos fundacionales de GenAI (como la serie GPT o Claude) son, en esencia, la cúspide evolutiva de NLP (Ahmad et al., 2023). La GenAI emplea modelos de lenguaje basados en la arquitectura Transformer, cuyo mecanismo de atención les permite ponderar el contexto para predecir el siguiente token más probable.

En los sistemas agénticos, el NLP va más allá de solo crear textos y se convierte en el modo de comunicación, control y razonamiento lógico (*ReAct* u orquestación multiagente) (IBM, 2025). El agente utiliza NLP para traducir la intención humana en comandos informáticos (por ejemplo, llamadas de bases de datos mediante APIs) y para "pensar en voz alta" en su módulo de auto-reflexión y toma de decisiones (Schmidgall et al., 2025)

En el ámbito educativo, el NLP ha permitido el desarrollo de herramientas como sistemas tutoriales inteligentes (*Intelligent Tutoring Systems*, ITS) que proporcionan retroalimentación personalizada en tiempo real (Sanabria-Navarro *et al.*, 2023). Asimismo, facilita la accesibilidad mediante la generación de contenido en múltiples idiomas y el análisis automatizado de la escritura de los estudiantes para detectar errores y sugerir mejoras. Sin embargo, su implementación plantea desafíos éticos, como la posible deshumanización de la enseñanza y la necesidad de mitigar sesgos algorítmicos que podrían afectar la equidad educativa (González-Sánchez et al., 2023).

A pesar de estos desafíos, el NLP sigue transformando la educación superior, ofreciendo soluciones innovadoras que democratizan el acceso al conocimiento y mejoran la interacción entre estudiantes y docentes mediante aplicaciones cada vez más avanzadas.

1. 4. 2. Computer Vision (CV)

La visión artificial o CV, permite a las máquinas interpretar y comprender información visual proveniente de imágenes, videos y datos tridimensionales.

Además, la CV combina algoritmos avanzados, matemáticas y procesamiento de imágenes para replicar la percepción visual humana. Las máquinas pueden identificar formas, colores y movimientos usando técnicas como la segmentación, el filtrado y el reconocimiento de objetos, lo que ayuda a comprender el entorno de manera automática (Kisaecanin et al.,

2009). Además, los avances en sistemas embebidos (*Embedded Systems*) han permitido incluir esta tecnología en dispositivos portátiles y aplicaciones de bajo consumo energético, extendiendo su uso a la vida cotidiana y el ámbito educativo.

Su objetivo principal es automatizar tareas que requieren percepción visual, como el reconocimiento de objetos, la segmentación de imágenes y la detección de patrones (Li *et al.*, 2020).

En educación, la CV tiene diversas aplicaciones. Por ejemplo, en el análisis de la participación estudiantil, puede evaluar expresiones faciales y lenguaje corporal para determinar niveles de atención en clases presenciales o virtuales. Estas herramientas permiten a los docentes ajustar su estrategia pedagógica en tiempo real para optimizar la retención de conocimientos (González-Sánchez *et al.*, 2023). En la supervisión de exámenes en línea, la CV facilita la detección de movimientos inusuales o intentos de fraude, asegurando la integridad de las evaluaciones (Li *et al.*, 2020).

Una de sus principales ventajas es la capacidad de procesar datos en tiempo real. Algoritmos optimizados y *hardware* eficiente permiten el monitoreo continuo de interacciones en el aula, como el uso correcto de materiales de laboratorio o la participación en experimentos científicos (Sanabria-Navarro *et al.*, 2023). En los laboratorios STEM, por ejemplo, la CV puede analizar el movimiento de partículas en entornos físicos o químicos, proporcionando un aprendizaje más visual y práctico.

Los agentes de los sistemas integran CV para supervisar el entorno áulico e interactuar de forma inmersiva. En entornos interactivos de realidad virtual (*Engage* o laboratorios robóticos autónomos), los agentes procesan en tiempo real datos biométricos de seguimiento visual (*gaze tracking*), postura y gestos corporales de los alumnos. Esto permite inferir sus niveles de atención o fatiga cognitiva (*embodied cognition*) y adaptar proactivamente la instrucción física o virtual (Burger *et al.*, 2020; IEEE, 2026).

Además, el CV ha mejorado significativamente la inclusión educativa. Aplicaciones como Microsoft Seeing AI y Be My Eyes utilizan esta tecnología para describir documentos impresos, gráficos y entornos físicos, facilitando el acceso a la información para usuarios con discapacidades visuales (Cornejo-Plaza & Cippitani, 2023). De esta manera, CV fomenta un aprendizaje accesible, asegurando que todos los estudiantes, independientemente de sus capacidades, puedan participar activamente en el proceso educativo.

La AIG, con el desarrollo de modelos multimodales (como GPT-4o o Gemini 1.5), puede interpretar diagramas, gráficos de rendimiento académico, ecuaciones escritas a mano y escaneos de documentos de admisión (University of Oxford, 2024; UPCEA, 2026).

1. 5. El ecosistema de datos en la Educación Superior. *Data Analytics*

La datificación de la educación ha generado una enorme cantidad de datos sobre el comportamiento de los estudiantes, su desempeño académico y la efectividad de las estrategias de enseñanza. Para gestionar esta información de manera eficiente, las instituciones educativas han integrado herramientas avanzadas de análisis de datos, como el *Big Data* (BD) y la Analítica de Datos (DA). Estas tecnologías permiten transformar vastos volúmenes de información en conocimiento aplicable, facilitando la toma de decisiones basada en evidencia y mejorando la experiencia de aprendizaje (Siemens & Gasevic, 2012).

1. 5. 1. *Big Data* y Analítica

El uso de datos en el ámbito educativo, particularmente en la educación superior, ha cobrado mayor relevancia en la última década. Aunque la generación de datos académicos ha sido una constante en las universidades desde el siglo XIX, lo que ha cambiado es el volumen, la velocidad y la variedad de estos datos, características clave del BD aplicado a la educación (Tsai *et al.*, 2020). Este proceso ha dado lugar a la denominada datificación del aprendizaje, donde cada dato almacenado puede revelar información clave sobre el comportamiento y resultados de los estudiantes (Jones, 2019).

El análisis de grandes volúmenes de datos, conocido como *Big Data for Education* (BDE) o ciencia de datos para la educación, ha permitido a las IES implementar estrategias para mejorar la experiencia de los actores educativos. Viberg *et al.* (2018) señalan que el primer gran reto tecnológico en este ámbito fue la incorporación de herramientas tecnológicas y de *Learning Management Systems* (LMS).

Si bien el BD en educación permite una mejor toma de decisiones basada en evidencia (Siemens & Gasevic, 2012), su uso también plantea desafíos éticos y de privacidad. La recopilación masiva de datos estudiantiles sin regulaciones claras podría comprometer la autonomía del estudiante, reforzando la necesidad de modelos que integren la privacidad por diseño y gobernanza de datos (Williamson, 2018). Como se explorará en el Capítulo 3, enfoques como el aprendizaje federado (Kairouz *et al.*, 2021) ofrecen soluciones para minimizar estos riesgos al descentralizar la información.

Desde principios de la década de 1980, se ha medido la combinación de IA, teorías del aprendizaje y prácticas educativas para mejorar los resultados de aprendizaje de los estudiantes usando computadoras (Rienties *et al.*, 2020). Como resultado, han surgido subcampos específicos que aplican la IA y el BD en la educación, entre los que destacan la *Educational Data Mining* (EDM) y los LA.

La IA agéntica puede operar sobre los flujos de datos de las instituciones educativas en tiempo real, ingiriendo continuamente información de sistemas como el SIS, el LMS y el CRM institucional (Higher Education Marketing, 2025; Microsoft, 2025). Un agente de soporte estudiantil requiere acceso a flujos masivos de datos —registros financieros, asistencia, calificaciones— para coordinar flujos de alerta eficaces.

El *Data Analytics* (DA) es el proceso sistemático de recopilar, organizar, procesar y analizar grandes volúmenes de información para identificar patrones, tendencias y relaciones significativas. Su propósito principal es transformar datos crudos en conocimientos prácticos que faciliten la toma de decisiones fundamentadas. En la educación, el DA permite transformar grandes volúmenes de información sobre el aprendizaje y desempeño de los estudiantes en conocimiento práctico y aplicable. En este ámbito, se concreta en los LA, los cuales se centran en la medición, recopilación y análisis de datos sobre los estudiantes y sus contextos con el objetivo de optimizar tanto el aprendizaje como los entornos educativos (Siemens, 2013).

El DA combina estadísticas, algoritmos de ML y tecnologías de bases de datos para proporcionar análisis avanzados en tres enfoques principales:

1. **Análítica descriptiva:** se centra en resumir y comprender lo que ocurrió en el pasado, como tendencias de desempeño académico o tasas de abandono escolar.
2. **Análítica predictiva:** utiliza modelos matemáticos y algoritmos para prever lo que podría suceder en el futuro, como la identificación de estudiantes en riesgo de deserción (Arnold & Pistilli, 2012; Baker & Siemens, 2014).
3. **Análítica prescriptiva:** sugiere acciones específicas basadas en los datos analizados, ayudando a optimizar procesos y mejorar resultados, como la personalización del aprendizaje o la evaluación del desempeño docente (Verbert et al., 2014).

El DA también se potencia al integrarse con otras áreas de la IA, como el NLP y CV. Por ejemplo, los datos textuales generados en foros educativos pueden analizarse con NLP para evaluar la participación y calidad de las discusiones. Paralelamente, las imágenes capturadas mediante reconocimiento facial se procesan con CV para medir la atención de los estudiantes.

La analítica combina estos datos, permitiendo a los docentes ajustar estrategias pedagógicas en tiempo real (Gaevi et al., 2015).

1. 5. 2. *Learning Analytics* (LA) y *Educational Data Mining* (EDM)

Las analíticas de aprendizaje o *Learning Analytics* (LA) han evolucionado como un campo interdisciplinario que combina BD, ML y estadística para analizar el comportamiento

estudiantil y optimizar los entornos educativos (Siemens & Gasevic, 2012). Su objetivo principal es identificar patrones en los datos generados por los estudiantes y utilizar esta información para mejorar los procesos de enseñanza y aprendizaje en la educación superior.

A nivel metodológico, los LA emplean diferentes técnicas de análisis de datos:

1. Modelado predictivo, que utiliza ML para anticipar el desempeño de los estudiantes y detectar posibles riesgos de deserción (Kotsiantis *et al.*, 2013).
2. Análisis de correlaciones, que permite identificar relaciones entre distintos factores educativos, como el tiempo de estudio y el rendimiento académico (Greller & Drachsler, 2012).
3. Agrupamiento, que segmenta a los estudiantes en grupos según su estilo de aprendizaje y nivel de desempeño, facilitando la personalización de estrategias pedagógicas (Romero & Ventura, 2010).

Las LA han sido fundamentales para la identificación de estudiantes en riesgo de deserción, el monitoreo del progreso estudiantil y la personalización del aprendizaje. Sin embargo, su implementación plantea desafíos como la privacidad de los datos, la equidad en el acceso a estas herramientas y la necesidad de mitigar sesgos algorítmicos en la toma de decisiones académicas (Williamson, 2018).

Uno de los casos más representativos es el de la *Open University* del Reino Unido, que ha desarrollado plataformas de LA para predecir el desempeño académico de los estudiantes. Gracias a estas herramientas, han logrado reducir en un 15% la tasa de abandono, permitiendo a los docentes intervenir de manera oportuna (Rienties *et al.*, 2020).

Otro ejemplo es *Course Signals*, un sistema desarrollado por la Universidad de Purdue, que emplea algoritmos de IA para monitorear el progreso de los estudiantes y generar alertas en caso de bajo rendimiento. Estudios sobre su implementación han demostrado que el uso de estas alertas ha incrementado la retención estudiantil en un 21% (Siemens & Baker, 2012). Además de la personalización del aprendizaje, los LA han optimizado la gestión educativa mediante:

1. Evaluación de estrategias pedagógicas, proporcionando datos en tiempo real sobre la efectividad de métodos de enseñanza (Ferguson, 2012).
2. Automatización de la retroalimentación, permitiendo que los estudiantes reciban respuestas inmediatas sobre su desempeño en actividades académicas.
3. Análisis de participación estudiantil, identificando patrones de interacción en plataformas educativas para mejorar el compromiso con el aprendizaje (Prinsloo, 2020).

Recientemente, los agentes de IA comienzan a transformar la analítica del aprendizaje pasiva en una analítica interactiva de carácter agéntico. Un agente dedicado a esta función puede ingerir continuamente datos de la interacción del alumno en el campus virtual —por ejemplo, la inactividad prolongada en el portal del curso— y señalar de forma proactiva indicios de desvinculación académica antes de que el personal lo advierta (Higher Education Marketing, 2025). En su desarrollo más avanzado, estos sistemas podrían llegar a reestructurar dinámicamente el material didáctico y aplicar intervenciones en tiempo real, en línea con la capacidad de los sistemas agénticos de monitorear su progreso y ajustar su comportamiento hacia metas definidas (Abou Ali et al., 2025).

A pesar de estos avances, las LA presentan desafíos significativos. La recopilación masiva de datos plantea preocupaciones sobre privacidad y ética en el manejo de información estudiantil. La necesidad de garantizar un uso transparente y responsable de los datos es clave para evitar decisiones algorítmicas sesgadas o la vigilancia excesiva del comportamiento estudiantil (Williamson, 2018). El monitoreo continuo de un estudiante por un sistema automático puede dar como resultado un *paternalismo algorítmico*, donde la máquina determina qué y cómo debe estudiar el alumno según predicciones, lo que afecta su libertad para formar su personalidad y su habilidad para adaptarse en la educación.

El *Educational Data Mining* (EDM) es un campo interdisciplinario que utiliza técnicas de IA, aprendizaje automático y estadística para analizar grandes volúmenes de datos educativos y descubrir patrones ocultos en el aprendizaje (Baker & Yacef, 2009). Su objetivo es proporcionar modelos predictivos que permitan mejorar la enseñanza y la experiencia de los estudiantes a través del análisis de datos masivos (Romero & Ventura, 2010).

Algunas de las principales aplicaciones de EDM incluyen:

1. Predicción del rendimiento estudiantil: modelos de ML que identifican estudiantes en riesgo de fracaso académico o abandono (Kotsiantis et al., 2013).
2. Personalización del aprendizaje: sistemas que adaptan los contenidos educativos a las necesidades específicas de cada estudiante en función de su desempeño previo (Krumm et al., 2014). La IA agéntica puede aprovechar los modelos predictivos refinados por la minería de datos educativos para regular sus flujos de trabajo (Romero y Ventura, 2020; Abou Ali et al., 2025). Mientras EDM estudia la estructura matemática del riesgo de abandono o la integración de un grupo, el agente de IA usa estos modelos de clasificación en su módulo de toma de decisiones para generar alertas institucionales adaptativas (Abou Ali et al., 2025).
3. Optimización de recursos educativos: análisis de tendencias en el uso de materiales digitales y plataformas de aprendizaje para mejorar su eficacia (Siemens & Baker, 2012).

La EDM permite a las instituciones educativas comprender cómo aprenden los estudiantes y mejorar la efectividad de sus metodologías pedagógicas. Sin embargo, su uso también requiere consideraciones éticas, como la transparencia en la interpretación de los datos y la protección de la privacidad estudiantil (Romero & Ventura, 2020).

El EDM y las LA son dos disciplinas complementarias que aplican técnicas de análisis de datos en entornos educativos. Aunque ambos campos comparten el objetivo de mejorar la enseñanza y el aprendizaje a través del análisis de datos, existen diferencias clave en sus enfoques y metodologías (Baker & Siemens, 2014).

1.6. Aplicaciones Prácticas de la IA en la Educación Superior

Como se ha venido mostrando, la IA está generando un impacto significativo en la educación superior. Su capacidad para analizar datos a gran escala, automatizar procesos y personalizar el aprendizaje ha facilitado el desarrollo de entornos académicos más eficientes y adaptativos (Siemens & Baker, 2012). La EDM explora la estructura de los datos para descubrir conocimiento, mientras LA lo aplica directamente en la enseñanza y la gestión educativa.

Entre las aplicaciones más comunes encontramos los *Learning Management Systems* (LMS), los *Intelligent Tutoring Systems*, (ITS) y la Automatización de Evaluaciones, cada uno de los cuales ha transformado la manera en que docentes y estudiantes interactúan con el conocimiento.

En los últimos años, estas aplicaciones se han visto potenciadas por la incorporación de herramientas basadas en GenAI y, de manera emergente, por sistemas de IA Agéntica. En nuestro país la Encuesta Nacional ENIAG 2025 (SEP, 2025) confirma la magnitud del fenómeno que esta investigación aborda: el 93% de los estudiantes y el 95% de los docentes conocen la IA generativa, y más del 60% de ambos grupos la utiliza de manera cotidiana, con un uso ligeramente mayor entre hombres (69.7%) que entre mujeres (62.7%).

La GenAI facilita la generación de contenidos educativos, materiales didácticos y retroalimentación personalizada dentro de los LMS. Por su parte, la IA Agéntica amplía las capacidades del docente mediante la planificación y ejecución autónoma y coordinada de tareas orientadas al logro de objetivos específicos.

Estas innovaciones están transformando progresivamente la manera en que las IES gestionan los procesos de enseñanza, aprendizaje y acompañamiento estudiantil (UNESCO, 2023; Abou Ali et al., 2025). A continuación, se presentan algunas de las principales aplicaciones prácticas.

1. 6. 1. *Learning Management Systems (LMS) y Tutoría Inteligente (ITS)*

Los sistemas de gestión de aprendizaje o LMS, son sistemas o plataformas tecnológicas diseñadas para administrar, entregar y monitorear actividades educativas. Estas herramientas proporcionan un entorno digital estructurado para la enseñanza y el aprendizaje, permitiendo a las instituciones educativas gestionar el contenido, las evaluaciones y las interacciones de los estudiantes. Además, recopilan datos valiosos, como asistencia, progreso en los módulos de aprendizaje y resultados de evaluaciones, que son fundamentales para la Analítica de Aprendizaje (LA) (Cornejo-Plaza & Cippitani, 2023).

La integración de la IA en los LMS ha mejorado significativamente su funcionalidad, permitiendo el análisis de datos en tiempo real para personalizar la experiencia de aprendizaje. Gracias al análisis de los datos de los estudiantes (LA), estos sistemas pueden identificar patrones de comportamiento estudiantil, predecir el desempeño académico y recomendar materiales específicos para fortalecer áreas de mejora (Prinsloo, 2020).

Los LMS también son esenciales para la implementación de modelos adaptativos de aprendizaje, ya que los datos recopilados se utilizan para ajustar el contenido y las actividades en función de las necesidades individuales de los estudiantes. Esto permite una experiencia de aprendizaje más personalizada y efectiva, maximizando las oportunidades de éxito académico. Por ejemplo, plataformas como Canvas pueden sugerir recursos adicionales o actividades remediales a estudiantes que presentan dificultades en temas específicos, basándose en predicciones generadas por herramientas de LA (Rojas & Chiappe, 2024).

Desde el punto de vista institucional, los LMS también desempeñan un papel clave en la toma de decisiones estratégicas. A través del análisis de datos, las universidades pueden evaluar la efectividad de sus programas educativos, mejorar la planificación curricular y optimizar la asignación de recursos. Esta capacidad analítica ha permitido la implementación de estrategias de intervención temprana para reducir tasas de abandono y mejorar la retención estudiantil (Sanabria-Navarro et al., 2023).

Más recientemente, diversos LMS han comenzado a incorporar funcionalidades basadas en GenAI, permitiendo la creación automática de contenidos, cuestionarios, actividades de aprendizaje y retroalimentación personalizada para estudiantes. En otras palabras, los profesores integran copilotos de GenAI en los LMS para simplificar la creación de contenidos educativos estructurados (rúbricas, syllabus, baterías de preguntas de examen) y para dar a los estudiantes retroalimentación textual enriquecida y automatizada de sus aportes escritos.

Asimismo, la evolución hacia sistemas de IA Agéntica abre la posibilidad de que estas plataformas no solo recomienden recursos educativos, sino que también coordinen acciones específicas, como programar actividades, monitorear avances académicos y generar alertas tempranas para docentes y administradores educativos (Chan, 2023; Abou Ali et al., 2025).

Un ejemplo de esta transición es la plataforma Canvas: su suite IgniteAI incorpora flujos de modelos de lenguaje dentro del LMS, captura evidencia del aprendizaje que se integra automáticamente al libro de calificaciones y, con el lanzamiento de IgniteAI Agent, orquesta flujos de trabajo de múltiples pasos —como la creación de módulos o el ajuste de plazos por grupos de estudiantes— a partir de instrucciones del docente (Instructure, 2025, 2026). Sin embargo, la aparición de agentes estudiantes que rinden exámenes de forma automatizada en milisegundos dentro del LMS introduce un severo riesgo de evasión de la evaluación auténtica (WCET, 2026).

Los llamados sistemas de tutoría inteligente o ITS han transformado la educación superior al proporcionar retroalimentación personalizada y adaptativa. A diferencia de los LMS, que funcionan principalmente como plataformas de gestión de contenido, los ITS utilizan técnicas avanzadas de ML y NLP para interactuar con los estudiantes de manera más dinámica y personalizada (Tomalá De la Cruz et al., 2023).

La interconexión entre estos sistemas —LMS y ITS— y el análisis de datos académicos (LA) es clave en la educación digital. Lo explicamos a continuación: los LMS proporcionan la infraestructura necesaria para la recopilación y gestión LA, estos datos son analizados por la IA para identificar patrones de aprendizaje. A partir de estos datos, los ITS pueden adaptar las experiencias educativas a las necesidades individuales de los estudiantes, optimizando así los procesos de enseñanza.

Uno de los primeros ITS fue SCHOLAR, desarrollado en la década de 1970, el cual ajustaba la enseñanza en función del conocimiento previo del estudiante. Desde entonces, la evolución de la IA ha permitido la creación de sistemas más sofisticados, como MATHia de Carnegie Learning o MyLab Math de Pearson, que adaptan los problemas matemáticos en tiempo real según el rendimiento del usuario (Mousavinasab et al., 2021).

Estos sistemas incorporan modelos de aprendizaje adaptativo que ajusta la dificultad de los ejercicios en función del progreso del estudiante. Además, el uso de chatbots basados en IA generativa permite a los alumnos resolver dudas en tiempo real, simulando la interacción con un tutor humano.

Además, el uso de chatbots basados en GenAI permite a los alumnos resolver dudas en tiempo real, generar explicaciones adaptadas a sus necesidades y recibir apoyo personalizado durante el proceso de aprendizaje (Kasneci et al., 2023). Más recientemente, la incorporación de capacidades agénticas en estos sistemas permite que algunos asistentes educativos no solo respondan preguntas, sino que también planifiquen actividades, den seguimiento a metas de aprendizaje y coordinen recursos educativos de manera más autónoma, fortaleciendo así el acompañamiento individualizado del estudiante (Bandi et al., 2025). El ITS agéntico no solo conversa con el alumno, sino que se convierte en un co-creador de aprendizaje activo de nivel colegiado (Williams, 2025). El tutor agéntico evalúa de forma autónoma cuándo aplicar

andamiajes restrictivos, diseña dinámicamente entornos interactivos basados en el desempeño, orquesta simulaciones de alta fidelidad, interactúa con tutores de otros alumnos para generar debates cooperativos cruzados, y recopila de manera proactiva evidencias del trabajo del estudiante para construir su portafolio de certificación académica final de forma automática (Williams, 2025).

Los estudios destacan que los estudiantes que usan estos sistemas experimentan mayores niveles de compromiso, mejor desempeño académico y una retención de conocimientos superior a la obtenida con métodos tradicionales (Kamalov et al., 2023).

1. 6. 2. Automatización de evaluaciones

La evaluación del desempeño estudiantil es un aspecto crucial en la educación superior, y la IA ha facilitado su automatización mediante herramientas que agilizan la calificación y retroalimentación de tareas.

En el ámbito de la evaluación escrita, los sistemas de IA han demostrado ser especialmente útiles en la corrección automática de ensayos y tareas. Modelos avanzados como BERT y GPT, basados en redes neuronales profundas, pueden analizar la estructura, coherencia y argumentación de textos, proporcionando comentarios detallados sobre la calidad de la escritura (Ramesh & Sanampudi, 2022).

La expansión de la GenAI ha fortalecido estos procesos mediante herramientas capaces de producir retroalimentación detallada, sugerencias de mejora y explicaciones adaptadas al nivel de desempeño de cada estudiante. Asimismo, el desarrollo de sistemas con capacidades agénticas podría permitir, en el futuro, la gestión integral de procesos de evaluación (UNESCO, 2023; Abou Ali et al., 2025).

Herramientas como Turnitin y Grammarly emplean NLP para evaluar gramática, estilo y plagio, permitiendo a los estudiantes mejorar sus habilidades de escritura mediante retroalimentación automatizada (Ramesh & Sanampudi, 2022; Ding & Zou, 2024). En asignaturas cuantitativas, los avances en CV han optimizado la corrección de exámenes en áreas como matemáticas y ciencias.

Plataformas como Gradescope, desarrollada en la Universidad de California, emplean algoritmos de reconocimiento óptico de caracteres (OCR) y aprendizaje profundo para analizar respuestas escritas a mano y calificarlas con un 70% de reducción en el tiempo de evaluación (Ramesh & Sanampudi, 2022).

Mediante el uso de NLP y CV, los sistemas automatizados pueden corregir exámenes, analizar ensayos y evaluar el desempeño de los estudiantes con rapidez y precisión (Ramesh & Sanampudi, 2022; Ding & Zou, 2024).

A medida que estas tecnologías evolucionan, su integración responsable en la educación superior permitirá mejorar los procesos de evaluación, optimizar la enseñanza y proporcionar retroalimentación más precisa y personalizada a los estudiantes.

Desde nuestro punto de vista, la implementación de estas requiere un equilibrio entre la automatización del aprendizaje y la interacción humana. La supervisión docente sigue siendo esencial para garantizar un acompañamiento adecuado, promoviendo un aprendizaje significativo y evitando una dependencia excesiva de la IA. Como señala Williamson (2018), es fundamental desarrollar marcos de regulación y supervisión docente para asegurar que la IA complemente, y no reemplace, el juicio pedagógico humano.

1. 6. 3. Gestión Educativa y Administrativa

Más allá del ámbito académico, la IA está optimizando la administración universitaria mediante la automatización de procesos, la mejora de la comunicación institucional y el fortalecimiento de la toma de decisiones basada en datos (Mayer-Schnberger & Cukier, 2014; Williamson, 2018).

Uno de los ámbitos donde esta transformación resulta más evidente es la atención al estudiantado. Herramientas conversacionales basadas en IA permiten gestionar procesos de admisión, inscripción, programación de cursos y servicios administrativos con mayor rapidez y disponibilidad. Un caso ampliamente documentado es Pounce, implementado en la Universidad Estatal de Georgia, cuyo uso contribuyó a reducir aproximadamente un 21% el abandono estudiantil al proporcionar asistencia inmediata durante trámites administrativos y resolver consultas frecuentes (Liu et al., 2023). Más recientemente, diversas instituciones han ampliado estas aplicaciones mediante asistentes inteligentes especializados para optimizar la comunicación institucional, la administración de la investigación y la gestión integral del ciclo de vida estudiantil.

En este contexto, la Universidad Rey Juan Carlos (España) implementó un proyecto piloto orientado a optimizar la comunicación entre el estudiantado y la secretaría académica del Máster Universitario en Formación del Profesorado. Para ello, desarrolló un modelo de atención en dos niveles que integró repositorios inteligentes contruidos con NotebookLM para la consulta autónoma de normativas, planes de estudio y reglamentos. Junto con un asistente interno denominado GemAI, diseñado para apoyar al personal administrativo en la elaboración de respuestas a las consultas recibidas por correo electrónico. Los resultados mostraron una disminución del 21.4% en el volumen global de correos electrónicos, así como reducciones superiores al 60% en consultas relacionadas con procedimientos de evaluación

y extensiones de asistencia. En contraste, aumentó la proporción de consultas complejas relacionadas con el Trabajo Fin de Máster, lo que evidencia que la automatización de tareas rutinarias permitió que el personal dedicara mayor tiempo a actividades que requieren acompañamiento y juicio profesional (Universidad Rey Juan Carlos, 2026).

Paralelamente, la IA ha fortalecido la gestión del ciclo de vida estudiantil mediante sistemas capaces de apoyar procesos de admisión, permanencia y éxito académico. En KL University (India), el agente conversacional de voz ARCA automatizó el seguimiento telefónico de aspirantes durante el proceso de admisión, ofreciendo atención en distintas lenguas y logrando incrementar aproximadamente 1.8 veces la asistencia a las jornadas de asignación de plazas. De manera complementaria, Illinois Institute of Technology desarrolló un sistema inteligente que integra información procedente del LMS, del sistema de información estudiantil, de plataformas financieras y de otros servicios institucionales para identificar tempranamente factores asociados al riesgo de abandono y generar alertas que permitan intervenciones oportunas dirigidas a estudiantes en situación de vulnerabilidad. Estas iniciativas se suman a experiencias ya establecidas, como Course Signals en la Universidad de Purdue, que utiliza modelos predictivos para identificar a estudiantes en riesgo académico y apoyar estrategias que ayudan a aumentar las tasas de retención estudiantil (Arnold & Pistilli, 2012).

La aplicación de la IA también comienza a extenderse hacia la gestión de la investigación. Un ejemplo es ORA gpt, desarrollado por la Office of Research Administration (ORA) de Emory University, como un asistente institucional para apoyar la administración de proyectos financiados con recursos externos. A diferencia de herramientas de propósito general, ORA gpt fue diseñado bajo un modelo de anclaje (*grounding*), conectándose exclusivamente con procedimientos operativos, normativas institucionales y documentación previamente validada por la universidad. Esta arquitectura permitió ofrecer asistencia permanente al personal responsable de la gestión de subvenciones. Además, automatizaron la elaboración de borradores de Procedimientos Operativos Estándar (SOP) y redujeron significativamente los tiempos de capacitación. Esto garantizó la protección de información confidencial y el cumplimiento de los estándares institucionales de ciberseguridad (Emory University, 2023).

La incorporación de asistentes inteligentes también comienza a extenderse a otras áreas de la gestión universitaria. La University of Tennessee Knoxville, ha documentado su utilización para apoyar procesos de recursos humanos mediante la elaboración de descripciones de puestos, la síntesis automática de cadenas extensas de correos electrónicos y otras tareas administrativas repetitivas. Esto permite que el personal concentre sus esfuerzos en actividades de mayor valor estratégico.

En conjunto, estas experiencias muestran que la IA ha dejado de limitarse al apoyo de los procesos de enseñanza y aprendizaje para convertirse en un componente estratégico de la

gestión universitaria. Su incorporación favorece la optimización de recursos, fortalece la comunicación institucional y mejora la toma de decisiones basada en evidencia.

1. 6. 4. Investigación Académica

La IA está revolucionando la investigación en educación superior al mejorar la recopilación de datos, el análisis de información y la difusión científica (Pedreño Muñoz et al., 2024; Liu et al., 2023). Entre sus principales aplicaciones destacan la búsqueda y síntesis de literatura científica, el análisis de grandes volúmenes de datos, el apoyo a la redacción académica y la gestión de proyectos de investigación.

La IA está transformando la forma en que los investigadores acceden, organizan y analizan la información científica. Desde hace varios años, plataformas como Google Scholar y Semantic Scholar emplean técnicas de procesamiento del lenguaje natural para analizar citas y recomendar publicaciones relevantes con mayor precisión. Más recientemente, herramientas como Consensus facilitan el acceso veloz a la evidencia científica mediante la síntesis de resultados de investigación, reduciendo considerablemente el tiempo destinado a la revisión bibliográfica. Asimismo, los avances recientes en modelos de lenguaje permiten elaborar resúmenes de literatura, comparar estudios, identificar vacíos de investigación y generar borradores iniciales de textos académicos, siempre bajo la supervisión crítica del investigador (Kasneci et al., 2023; UNESCO, 2023).

Además de facilitar el acceso al conocimiento científico, la IA comienza a desempeñar un papel relevante en la recuperación y estructuración de información histórica para la investigación. Un ejemplo es el trabajo desarrollado por Charles Godfray, de la Oxford Martin School, quien empleó un modelo de IA para analizar más de 500 páginas de literatura científica histórica sobre ecología de comunidades y transformar información no estructurada en bases de datos organizadas para su análisis. Este tipo de aplicaciones demuestra el potencial de estas tecnologías para recuperar conocimiento científico previamente inaccesible y acelerar significativamente procesos que tradicionalmente requerían semanas de trabajo manual (University of Oxford, 2024).

La evolución reciente de estas tecnologías también está transformando la ejecución misma de la investigación científica. Los denominados *Self-Driving Laboratories* (SDL) integran algoritmos de IA con plataformas robóticas capaces de diseñar, ejecutar y optimizar experimentos de manera autónoma. Estos laboratorios automatizan la formulación de hipótesis, la planificación experimental, la ejecución de pruebas y el análisis iterativo de resultados, acelerando los ciclos de experimentación, optimizando el uso de recursos y favoreciendo el descubrimiento científico mediante procesos continuos de aprendizaje y experimentación automatizada (Canty et al., 2025). De manera complementaria, comienzan a surgir plataformas multiagente capaces de colaborar con los investigadores durante distintas fases del proceso científico, desde la revisión bibliográfica y la ejecución de análisis computacionales hasta la elaboración de reportes de investigación. Un ejemplo es Agent

Laboratory, que emplea agentes basados en modelos de lenguaje para coordinar tareas de investigación, planificar experimentos, analizar resultados y generar documentación científica, ampliando las posibilidades de colaboración entre investigadores y sistemas inteligentes (Schmidgall et al., 2025).

La IA también ha fortalecido los procesos de redacción y publicación científica. Plataformas como Writefull y Grammarly contribuyen a mejorar la gramática, la claridad y la coherencia de los manuscritos académicos, mientras que las herramientas de traducción automática facilitan la comunicación científica entre investigadores de distintos contextos lingüísticos. Asimismo, algunos sistemas permiten identificar revistas potencialmente adecuadas para la publicación de resultados de investigación y optimizar la preparación de los manuscritos conforme a los requisitos editoriales (Farber, 2025).

En conjunto, estas aplicaciones evidencian que la IA está redefiniendo la investigación universitaria al automatizar tareas repetitivas, facilitar el acceso al conocimiento científico y fortalecer la gestión de proyectos de investigación. Su capacidad para procesar grandes volúmenes de información y apoyar distintas etapas del trabajo científico permite optimizar tiempos y recursos, sin sustituir el juicio crítico ni la responsabilidad del investigador (Zawacki-Richter et al., 2019).

No obstante, la incorporación de estas tecnologías también plantea desafíos que deben abordarse desde una perspectiva ética y jurídica. La transformación de los procesos de investigación requiere mantener la integridad científica, asegurar que se use estas herramientas de manera transparente y considerar las consecuencias del cambio en los roles tradicionales de los docentes e investigadores (Sanabria-Navarro et al., 2023; Comisión Europea, 2020).

Uno de los aspectos que requiere mayor atención es la protección de los datos utilizados durante los procesos de investigación y gestión académica. A medida que las IES emplean volúmenes crecientes de información para apoyar decisiones pedagógicas, administrativas y científicas, resulta indispensable garantizar la privacidad de los datos personales y asegurar que su tratamiento se realice bajo principios de licitud, seguridad y consentimiento informado (Liu et al., 2023).

Por ello, resulta pertinente analizar cómo proteger, desde una perspectiva ética y jurídica, la privacidad de los estudiantes en el contexto específico de las IES en México. Esta problemática se desarrolla en los capítulos siguientes: el segundo aborda los principios éticos que orientan el uso responsable de la IA, el tercero analiza la figura jurídica del consentimiento informado y el cuarto presenta la propuesta que constituye la principal aportación de esta investigación.

1. 7. Riesgos y Desafíos Éticos

La creciente incorporación de sistemas de GenAI y, más recientemente, de IA Agéntica, ha ampliado tanto las oportunidades como los desafíos asociados al uso de la IA en la educación superior.

Los riesgos de la GenAI se suelen asociar con la creación de información incorrecta, sesgada o difícil de comprobar; los sistemas agénticos presentan problemas extras por su habilidad para planificar, coordinar y realizar acciones de manera autónoma, sin necesidad de supervisión humana (UNESCO, 2023; Abou Ali et al., 2025). A continuación, se exponen los riesgos y desafíos éticos considerados como principales.

1.7.1. Acceso Equitativo

La distribución equitativa del acceso a educación e investigación sobre IA es una preocupación destacada por la UNESCO (2021). Mientras países como China y Estados Unidos lideran en presencia y desarrollo de IAED, regiones como el África subsahariana, partes de Asia central y del sur y algunas zonas de América Latina aún no alcanzan el mismo nivel de avance (Pedreño Muñoz, 2024). Por su parte, esta distribución desigual se refleja en la correlación entre el PIB y la cantidad de publicaciones académicas sobre IAED, aunque hay excepciones notables (Liu et al., 2023).

Este riesgo implica que se vea afectado el derecho a una educación accesible, así como el principio de igualdad y no discriminación. La falta de acceso a tecnología de alta calidad puede dejar en desventaja a ciertos sectores del estudiantado.

Las IES deben adoptar políticas institucionales y medidas de apoyo que aseguren oportunidades de aprendizaje comparables, con independencia de la situación socioeconómica (Liu et al., 2023). Por su parte, tales medidas contribuyen a mitigar las desigualdades y a garantizar que cada estudiante tenga la posibilidad de beneficiarse de los avances tecnológicos en educación (Cordón, 2023). Aunado a lo anterior, consideramos que es oportuno llevar a cabo evaluaciones periódicas de brechas de acceso y monitoreo de equidad en el uso de herramientas de IAED.

1.7.2. Sesgos y Discriminación

La implementación de sistemas de IAED puede incorporar sesgos en los datos de entrenamiento o en los algoritmos, lo que podría derivar en discriminación contra determinados grupos de estudiantes, por ejemplo, por origen étnico o condición socioeconómica (Cordón, 2023). Por ejemplo, si se toman decisiones de admisión basadas en IAED, la existencia de sesgos puede traducirse en una reducción injustificada de becas u oportunidades para ciertos grupos de estudiantes (Liu et al., 2023). Otro ejemplo de sesgo se

puede ver en los ITS y en otros modelos que personalizan el aprendizaje. Estos pueden verse afectados por sesgos ocultos en los datos que utilizan, lo que puede llevar a resultados diferentes para estudiantes de diversos orígenes (Prinsloo, 2020).

En el caso de la GenAI, los prejuicios en los datos de entrenamiento pueden aparecer en contenidos, recomendaciones o respuestas que muestren estereotipos o representaciones desiguales. Los sistemas agénticos pueden aumentar el efecto de estos sesgos cuando las recomendaciones que generan se convierten en acciones o decisiones automáticas en procesos académicos o administrativos. Esto puede afectar la equidad y la inclusión en las instituciones educativas (Abou Ali et al., 2025).

Cuando hay sesgos y discriminación, se ven afectados el derecho a la igualdad y no discriminación, así como las garantías de imparcialidad en decisiones educativas automatizadas que son ejecutadas por los algoritmos de IA.

Por lo anterior, resulta necesario crear mecanismos y políticas institucionales para asegurar que las decisiones apoyadas por algoritmos en contextos educativos no discriminen a determinados grupos (Alonso-Rodríguez, 2024). Además, proponemos documentar criterios, asegurar supervisión humana, y habilitar mecanismos de explicación, impugnación y corrección.

1.7.3. Calidad Educativa

La automatización y personalización a través de la IAED pueden impactar la calidad de la educación (Cordón, 2023). Por una parte, la automatización de ciertos procesos puede liberar tiempo para que el profesorado se enfoque en actividades más atractivas y enriquecedoras con el estudiantado, y la IAED puede adaptar materiales y actividades educativos según las necesidades del estudiantado. Por su parte, existe el riesgo de que la sobre dependencia de la IAED para la enseñanza disminuya la interacción humana y la retroalimentación personalizada, lo que podría afectar negativamente la calidad del aprendizaje y la relación humana en el proceso educativo (González-Campos et al., 2024).

El uso cada vez mayor de herramientas basadas en GenAI puede fomentar la dependencia tecnológica cuando se utilizan en sustitución del pensamiento crítico, la creatividad, la resolución de problemas o la capacidad de generar conocimiento de manera autónoma por parte de los estudiantes. Además, la facilidad que ofrecen estas herramientas para crear textos, ensayos, presentaciones, código u otros contenidos de manera automática puede fomentar un aprendizaje superficial. Esto puede llevar a los estudiantes a enfocarse más en obtener respuestas rápidas que en reflexionar, analizar y construir su conocimiento. De manera similar, la incorporación de sistemas de IA Agéntica exige especial atención para evitar una delegación excesiva de decisiones pedagógicas o académicas que tradicionalmente corresponden a docentes y autoridades educativas. En ambos casos, la supervisión humana

resulta indispensable para preservar la calidad educativa y garantizar que la tecnología complemente, y no sustituya, los procesos formativos (UNESCO, 2023; Bandi et al., 2025).

Además, el uso intensivo de datos de comportamiento estudiantil puede conducir a una cultura de observación constante, lo que podría afectar negativamente el bienestar emocional y la autonomía (Prinsloo, 2020).

De modo sintético, estos riesgos pueden afectar derechos y principios fundamentales como la autonomía personal, la libertad académica, la integridad académica, la protección de los derechos de autor, la proporcionalidad en el uso de tecnologías y, en última instancia, la dignidad de los estudiantes. Es recomendable que las IES delimiten las finalidades pedagógicas del uso de IAED, eviten una vigilancia excesiva, informen de forma clara sobre el tratamiento de datos personales y aseguren la supervisión humana.

1.7.4. Privacidad y Seguridad de Datos

El uso no regulado por principios éticos y normativas claras puede traer como consecuencia la violación del derecho a la privacidad y de la protección de datos personales, confidencialidad y seguridad. La recopilación y procesamiento de datos personales del estudiantado plantean preocupaciones significativas en términos de privacidad y seguridad (Cordón, 2023). La implementación de LA agrega capas adicionales de complejidad (Marín, 2023) que revisaremos más adelante.

La recolección de datos detallados sobre el comportamiento del estudiantado puede incrementar la exposición de información personal sensible, lo que subraya la importancia de salvaguardar adecuadamente la privacidad y la seguridad de los datos en entornos educativos impulsados por la IA.

La expansión de herramientas basadas en GenAI ha incrementado la recopilación y procesamiento de información académica mediante interacciones continuas entre estudiantes, docentes y plataformas digitales. A ello se suma la aparición de sistemas de IA Agéntica que puede acceder, integrar y utilizar información proveniente de múltiples fuentes institucionales para ejecutar tareas o formular recomendaciones. Esta capacidad incrementa la necesidad de establecer mecanismos robustos de protección de datos personales, consentimiento informado y supervisión institucional sobre el uso de la información estudiantil (UNESCO, 2023; Bandi et al. 2025).

Por todo lo anterior, es necesario considerar marcos éticos y regulatorios que garanticen la protección efectiva de los datos, el consentimiento informado y fomenten una cultura de confianza en las IES que utilizan IA, especialmente la GenAI y la Agéntica.

CAPÍTULO 2. Principios Éticos de la IA y su Marco de Aplicación en Instituciones de Educación Superior.

En cumplimiento con el segundo objetivo específico de esta investigación, este capítulo examina y desarrolla los principios éticos que deben orientar la integración de la IA en la educación superior. Asimismo, presenta marcos de evaluación y gestión de riesgos que permiten traducir dichos principios en criterios operativos y verificables. Es imperativo que la comunidad educativa, en colaboración con expertos en tecnología y ética, forje un camino que aproveche el potencial de la IA al tiempo que salvaguarda los derechos fundamentales de privacidad y promueve una experiencia educativa equitativa e inclusiva (*U. S. Department of Education, Office of Educational Technology, 2023*).

Los principios éticos desarrollados en este capítulo son aplicables tanto a la IA Generativa (GenAI) como a la IA Agéntica. Sin embargo, ambas tecnologías presentan desafíos diferenciados. Mientras que la GenAI se orienta principalmente a la generación de contenidos, información y recomendaciones, la Agéntica incorpora capacidades de planificación, coordinación y ejecución autónoma de acciones orientadas a objetivos específicos. Por lo antes dicho, los principios éticos adquieren especial relevancia para garantizar que tanto los contenidos generados como las acciones ejecutadas por estos sistemas respeten los derechos fundamentales, la dignidad humana y los fines educativos de las instituciones de educación superior.

La implementación y el uso de la GenAI y agéntica, no deben limitarse a su aspecto técnico, sino que deben abordarse desde una perspectiva holística que contemple sus implicaciones sociales, éticas y culturales.

¿Cómo aproximarse a esta perspectiva holística de la ética en la IA? Siguiendo a Floridi (2013), se considera que las tecnologías emergentes deben diseñarse para respetar los valores humanos fundamentales, y por lo tanto estar al servicio de la persona.

2. 1. La IA como una Herramienta al Servicio de la Persona

La Pontificia Academia para la Vida (PAV) en su iniciativa *Rome Call for AI Ethics* (2020) enfatiza que la IA debe diseñarse para complementar la inteligencia humana, evitando que se convierta en un sustituto del razonamiento o la creatividad. El organismo señala que el principal riesgo consiste en que el uso de la IA reemplace la capacidad de las personas para desarrollar pensamiento crítico y toma de decisiones autónomas.

Aplicado a la educación, el organismo precisó que el riesgo es que la IA transforme la educación en un ejercicio mecánico, donde las respuestas sean generadas automáticamente sin que los estudiantes desarrollen habilidades analíticas propias (Pontificia Academia para la Vida, 2020). Si los estudiantes confían ciegamente en la IA para obtener respuestas, en

lugar de desarrollar sus propias estrategias de resolución de problemas, la educación podría perder su función formativa y convertirse en un proceso de dependencia tecnológica.

Por ello, la IA debe diseñarse para potenciar el pensamiento humano, funcionando como una herramienta que fomente la autonomía del aprendizaje, en lugar de un sistema que lo reemplace (Pontificia Academia para la Vida, 2020). En esta misma línea, León XIV (2026) advierte que la IA constituye una herramienta valiosa, pero requiere vigilancia permanente para evitar que la lógica tecnológica sustituya el discernimiento moral de las personas. La responsabilidad sobre sus decisiones nunca puede desplazarse completamente a los sistemas automatizados, pues corresponde siempre al ser humano orientar su desarrollo y utilización.

2. 1. 1. El Triángulo de la Irreductibilidad Humana

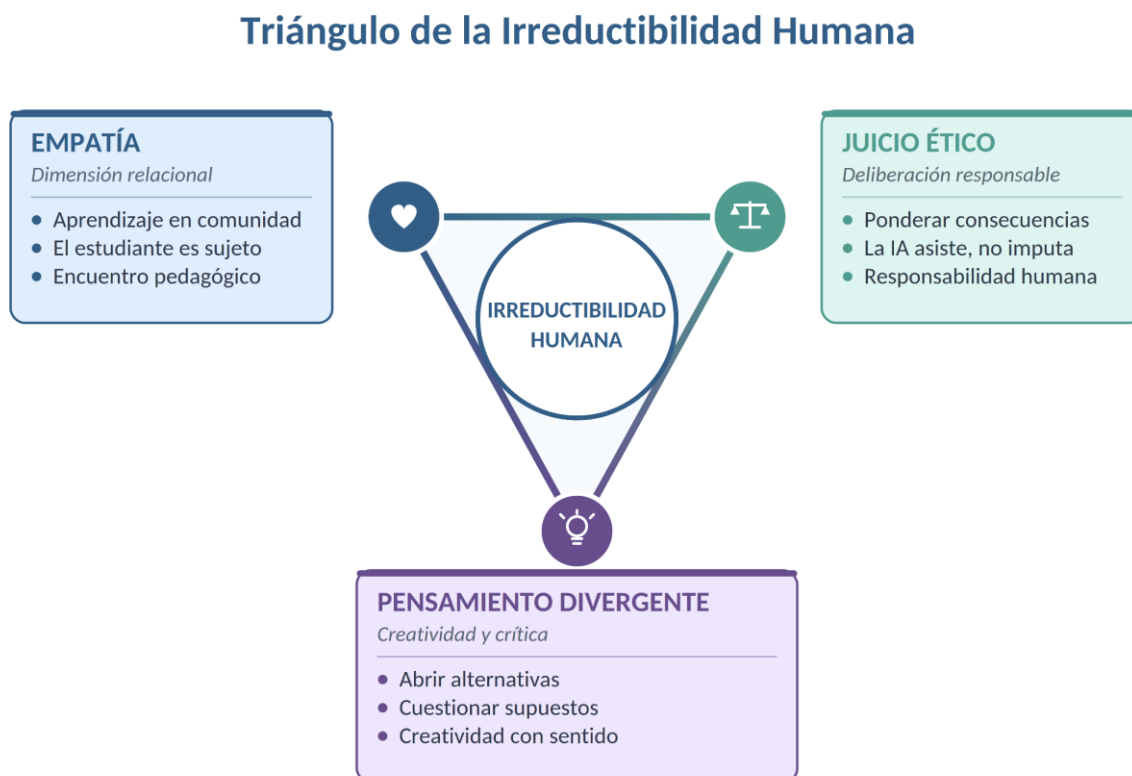
En el marco de esta investigación, la evaluación ética del uso de IA en educación superior requiere un criterio que vaya más allá de la eficiencia, la precisión o la mera conformidad normativa. Por ello, se propone el Triángulo de la Irreductibilidad Humana como un criterio antropológico-normativo. El triángulo que proponemos se apoya en tres capacidades constitutivas del ser humano y, por lo tanto, necesarias en la vida universitaria: empatía, juicio ético y pensamiento divergente. A continuación, desarrollaremos, brevemente, cada elemento.

La empatía remite a la dimensión relacional del aprendizaje. La universidad forma en comunidad y produce conocimiento en interacción. Por ello, el diseño y uso de IA debe evitar la cosificación del estudiantado como “perfil” o “caso” y debe preservar espacios de encuentro pedagógico donde el otro sea reconocido como sujeto y no solo como fuente de datos (Stein, 2005).

El juicio ético se refiere a la deliberación prudencial y responsable en contextos concretos. En educación superior, el juicio del docente y del estudiante se ejerce al interpretar información, ponderar consecuencias y decidir ante dilemas. La IA puede apoyar la decisión, pero no puede reemplazar la imputabilidad ni la responsabilidad propias del agente humano, especialmente cuando están en juego derechos, trayectorias académicas o bienestar (Aristóteles, 2000; Kant, 2003).

Por último, el pensamiento divergente remite a la capacidad de abrir alternativas, cuestionar supuestos y comenzar lo nuevo. En la universidad, esta dimensión se expresa en investigación, argumentación y creatividad con sentido. Aunque los sistemas generativos producen variaciones que parecen razonables, la educación universitaria necesita procesos de exploración, justificación y atención al significado que fomentan la innovación y el pensamiento crítico (Arendt, 1993; Frankl, 2004).

Figura 4. Triángulo de la Irreductibilidad Humana.



Nota. Elaboración propia con apoyo de Claude Sonnet 5 (Anthropic, 2026), mediante un proceso iterativo iniciado con el prompt "A partir de este texto, genera una infografía que refleje el triángulo de la irreductibilidad humana," y refinado mediante instrucciones sucesivas de la autora para el diseño gráfico, a partir del contenido y la estructura desarrollados por la autora.

El triángulo refuerza la idea de que la formación universitaria se debe orientar a desarrollar sujetos relacionales, morales y creativos. Por lo anterior, la pregunta decisiva no es si un sistema de IA puede imitar conductas, sino si su integración preserva y fortalece las capacidades que sostienen la dignidad, la agencia y el sentido del aprendizaje (Floridi, 2013; Pontificia Academia para la Vida, 2020).

Esta visión coincide con la reflexión más reciente del Magisterio de la Iglesia. En la encíclica *Magnifica Humanitas* (2026), León XIV sostiene que el desarrollo de la IA debe preservar siempre la primacía de la persona humana. De modo que sea la inteligencia humana, guiada por la conciencia y la libertad, la que determine responsablemente el uso y los límites de estas tecnologías. De este modo, la innovación tecnológica encuentra su legitimidad cuando permanece orientada al bien común y al respeto de la dignidad humana.

2. 2. Principios de Uso Ético de la IA en la Educación

Diversos autores (Deci & Ryan, 2000; Holmes et al., 2022; Nguyen et al., 2023) y organismos internacionales (UNESCO, 2019, 2021; OCDE, 2021) han identificado principios clave que deben guiar la implementación de la IA en la educación.

En general, en estos principios destacan la importancia de una gobernanza responsable, la protección de la autonomía y privacidad de los estudiantes y la necesidad de diseñar sistemas que empoderen tanto a estudiantes como a docentes en su proceso de enseñanza-aprendizaje.

El punto de partida compartido por los estudiosos del tema y las grandes agencias internacionales para elaborar sus principios es que la IA educativa solo es éticamente admisible si preserva la autonomía, la privacidad y la creatividad de quienes aprenden (Deci & Ryan, 2000, 2021; OCDE, 2021).

Un marco ético de referencia a nivel global es el *Rome Call for AI Ethics* (Pontificia Academia para la Vida, 2020) respaldado por más de un centenar de organizaciones como Microsoft, IBM y la FAO. En este marco se destaca que una IA ética debe proteger la libertad del ser humano y ponerlo en el centro de todo avance tecnológico.

Sobre esa base, y tomando como referencia el Marco General de Uso Ético de la IA de la Universidad Panamericana (Universidad Panamericana, 2025), se presentan a continuación los principios que guiarán el resto del capítulo.

2. 2. 1. Centrada en la Persona

El primer pilar ético, que de manera reiterativa se encuentra en la academia y los organismos internacionales, es que la IA debe estar centrada en la persona humana. Esto implica que los modelos educativos impulsados por IA deben ser diseñados para fortalecer el aprendizaje autónomo, fomentar la creatividad y garantizar que la tecnología esté al servicio de los valores humanos.

El principio de un sistema de IA centrado en el ser humano es el eje rector de cualquier marco ético que aspire a una implementación responsable de la tecnología en contextos educativos. Este principio establece que la IA debe ser diseñada, desarrollada e implementada para potenciar las capacidades humanas, sin sustituir la agencia, la autonomía ni la toma de decisiones críticas de los individuos involucrados. En educación superior, esto se traduce en sistemas que acompañan y amplifican el proceso formativo, pero que no desplazan la relación pedagógica ni estandarizan el aprendizaje de manera mecánica.

Este enfoque centrado en la persona se operacionaliza, en esta investigación, mediante el Triángulo de la Irreducibilidad Humana: una IA educativa es aceptable cuando fortalece (i) la relacionalidad del aprendizaje (empatía), (ii) la agencia moral y la responsabilidad (juicio ético) y (iii) la creatividad crítica y la exploración (pensamiento divergente), en lugar de sustituirlas o debilitarlas.

La UNESCO (2021) plantea que los sistemas de IA deben contribuir al desarrollo humano y social, fortaleciendo el aprendizaje autónomo y creativo. En ese sentido, una IA centrada en el ser humano no puede entenderse únicamente desde una perspectiva funcional, sino también

desde una dimensión formativa, ética y cultural, que respete la dignidad, la justicia y los derechos fundamentales de estudiantes, docentes y personal académico. Esto incluye rechazar enfoques deterministas o tecnocráticos, en los que la IA se impone como criterio único para la toma de decisiones educativas.

Desde la teoría de la autodeterminación, Deci y Ryan (2000) sostienen que el aprendizaje significativo se apoya en el respeto a la autonomía, la competencia y la relación, condiciones que deben guiar también el diseño de herramientas inteligentes. Si los sistemas de IA no reconocen estas necesidades psicológicas básicas, corren el riesgo de convertirse en tecnologías de control más que de empoderamiento.

El Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019 de la Comisión Europea, en el documento titulado *Directrices Éticas para una IA Confiable* (2019), coincide en que el enfoque centrado en el ser humano debe estar presente en todo el ciclo de vida del sistema de IA: desde la concepción, pasando por el diseño, la implementación, la evaluación y la supervisión. En esta visión, el ser humano no es un usuario pasivo, sino un agente activo con derecho a comprender, auditar y decidir cómo interactuar con la tecnología. Este principio adquiere especial relevancia en la educación superior, donde los estudiantes están formando sus competencias críticas, éticas y ciudadanas para la vida adulta.

En el ámbito académico, autores como Nguyen et al. (2023) advierten sobre los riesgos de la manipulación cognitiva por parte de sistemas de IA, especialmente si están entrenados para maximizar eficiencia sin atender al impacto pedagógico o emocional en los estudiantes. Por ello, los sistemas deben garantizar la posibilidad de cuestionamiento, revisión y retroalimentación humana, de modo que la tecnología no se convierta en un sustituto del juicio ético ni de la deliberación educativa.

Además, desde una perspectiva institucional, el Departamento de Educación de EE. UU. (2023) refuerza este enfoque, señalando que las decisiones educativas deben seguir lideradas por humanos. En esta línea, la centralidad humana también se refuerza mediante los requisitos de supervisión humana y rendición de cuentas del enfoque de IA fiable de la UE (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019). Este principio se relaciona con el respeto por la dignidad humana, la autonomía y libre desarrollo de la personalidad.

Por último, una IA centrada en la persona es tanto una cuestión técnica, como política y filosófica. Implica reconocer que el diseño tecnológico no es neutral, y que toda herramienta encarna una visión del mundo, del aprendizaje y de la relación entre poder y conocimiento. Por lo tanto, defender este principio implica que las universidades adopten una visión crítica sobre las soluciones tecnológicas, se involucren en su desarrollo y creen sistemas de gobernanza ética que aseguren un uso justo, transparente y transformador.

Este principio resulta especialmente relevante frente al crecimiento de la GenAI y de la IA Agéntica. Mientras la primera puede influir en la manera en que estudiantes y docentes producen conocimiento, la segunda puede intervenir en procesos de decisión y gestión educativa. En ambos casos, la tecnología debe permanecer subordinada al desarrollo integral de la persona y nunca sustituir el juicio humano.

2. 2. 2. Privacidad y Seguridad de los Datos

La recopilación y procesamiento de datos personales del estudiantado plantean preocupaciones significativas en términos de privacidad y seguridad (Cordón, 2023). La privacidad y la protección de datos personales constituyen uno de los pilares más sensibles y fundamentales en la implementación ética de la IA en la educación superior. En este contexto, es una exigencia legal, y al mismo tiempo un compromiso ético con la autonomía, la dignidad y la seguridad de las personas, particularmente cuando los datos recolectados incluyen información académica, socioeconómica, emocional o conductual de estudiantes y docentes.

En esa línea, Sedenberg y Hoffmann (2016) argumentan que el consentimiento informado no debe abandonarse ante las prácticas intensivas en datos, pues constituye el mecanismo histórico que operacionaliza el respeto por las personas y su capacidad de decidir sobre la información que les concierne. Esto resulta particularmente importante en educación, donde la recopilación masiva de datos puede tener efectos duraderos sobre la identidad digital y las oportunidades académicas de los estudiantes.

La UNESCO (2021) refuerza esta visión al establecer que los sistemas de IA deben garantizar la seguridad de los datos personales mediante encriptación robusta, anonimización y supervisión humana significativa. Asimismo, deben evitar prácticas de vigilancia invasiva o recolección excesiva que puedan vulnerar la libertad académica, el derecho al error o la privacidad cognitiva.

Asimismo, las Directrices Éticas para una IA fiable Grupo de Expertos de Alto Nivel de la Comisión Europea sobre IA (2019) y el documento titulado *AI Risk Management Framework (AI RMF)* de la *National Institute of Standards and Technology* (NIST) de EE. UU. (NIST, 2023) coinciden en que la evaluación responsable de un sistema no se agota en su desempeño técnico, sino que exige considerar impactos, gobernanza y controles durante todo el ciclo de vida. En consecuencia, la responsabilidad institucional se operacionaliza mediante directrices, herramientas y puntos de control. Por ejemplo, evaluaciones de impacto, registros de decisiones, protocolos de consentimiento informado y mecanismos de supervisión humana permiten documentar, revisar prácticas y corregir desviaciones de forma iterativa (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023).

Garantizar la privacidad de los datos de los estudiantes es una obligación de las IES, éstas deben cuidar en todo momento que los datos del alumnado se recaben, almacenen, procesen

y usen cuidando su privacidad. Consideramos clave que durante la primera etapa, la recabación, se solicite su consentimiento. Este consentimiento debe ser informado. Autores como Nguyen et al. (2023) subrayan que el consentimiento informado debe ser comprensible, revocable y específico, especialmente en contextos de aprendizaje automatizado, evitando la opacidad algorítmica.

Desde una visión sistémica, la privacidad y la protección de datos se incorporan como requisito de IA fiable y como riesgo a gestionar transversalmente. El enfoque de la UE incluye la privacidad y gobernanza de datos, mientras que el AI RMF del NIST exige mapear, medir y gestionar riesgos asociados a datos, sesgos y usos indebidos (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023). Esto implica que la protección de datos no puede depender solo de protocolos técnicos, sino que debe estar acompañada de una cultura institucional que promueva la alfabetización en derechos digitales, el empoderamiento del usuario y la evaluación constante de riesgos emergentes.

En cuanto a las estrategias técnicas, enfoques como la privacidad diferencial (Dwork & Roth, 2014) y el aprendizaje federado (Kairouz et al., 2021) ofrecen soluciones prometedoras para mitigar los riesgos de exposición de datos. La privacidad diferencial permite realizar análisis sin identificar directamente a los individuos, mientras que el aprendizaje federado entrena modelos de IA sin necesidad de centralizar la información sensible en un único servidor. Aunado a lo anterior, es importante considerar las técnicas de ciberseguridad para cuidar los datos.

Por todo lo anterior, garantizar la privacidad y la ciberseguridad de los datos en entornos de IA educativa requiere más que cumplir con una normativa: exige una visión ética comprometida con el respeto a los derechos digitales, la protección de la integridad estudiantil y la construcción de confianza entre usuarios, instituciones y tecnologías. La privacidad no debe ser tratada como un obstáculo al desarrollo de la IA, sino como su condición de posibilidad legítima y sostenible en el contexto de la educación superior.

En el caso de la GenAI, los riesgos se relacionan principalmente con el tratamiento de la información utilizada para generar contenidos y respuestas sin consentimiento del usuario. En los sistemas de IA Agéntica, el reto crece porque pueden acceder, juntar y usar información de varias fuentes para llevar a cabo acciones o hacer recomendaciones. Esto requiere mejores protecciones de datos, consentimiento informado y supervisión de las instituciones.

2. 2. 3. Transparencia y Explicabilidad

La transparencia constituye uno de los principios fundamentales para el desarrollo y uso ético de sistemas de IA confiables. Dentro de este principio, la explicabilidad ocupa un lugar central, ya que permite comprender cómo y por qué un sistema de IA genera una determinada

decisión o recomendación. En este sentido, ambos conceptos se encuentran estrechamente vinculados: mientras la transparencia busca que el funcionamiento y los procesos de los sistemas sean comprensibles y accesibles para los usuarios, la explicabilidad proporciona los elementos necesarios para interpretar el razonamiento seguido por el sistema y justificar sus resultados.

El Grupo de Expertos de Alto Nivel de la Comisión Europea establece que la transparencia debe aplicarse en tres niveles: en el diseño del sistema (trazabilidad del proceso), en el funcionamiento (explicabilidad del algoritmo)¹ y en la interacción con los usuarios (comunicación clara y comprensible). En el ámbito educativo, esto implica que tanto docentes como estudiantes deben saber qué tipo de datos se recopilan, con qué propósito, cómo se procesan y de qué manera se toman decisiones automatizadas (Grupo de Expertos de Alto Nivel sobre IA, 2019).

La explicabilidad (*explainability*) es uno de los componentes esenciales del principio de transparencia y resulta indispensable para lograr una IA confiable, auditable y ética, especialmente en entornos educativos donde las decisiones automatizadas pueden tener impactos directos en la trayectoria académica, emocional o social de los estudiantes. Este componente establece que los usuarios deben poder comprender cómo y por qué un sistema de IA ha llegado a una determinada decisión o recomendación, así como tener la posibilidad de cuestionarla, interpretarla y, cuando sea necesario, rechazarla o corregirla (Martínez, 2024).

Desde la perspectiva del Triángulo de la Irreductibilidad Humana (TIH), la explicabilidad es necesaria para favorecer el juicio ético: sin comprensión suficiente del proceso algorítmico, no existe deliberación informada ni posibilidad real de impugnar decisiones que afecten trayectorias educativas. Esto conecta con la crítica a las “cajas negras” en educación superior, donde la autoridad se expresa algorítmicamente y puede excluir el cuestionamiento del usuario si no se abren canales de interpretación y revisión (Prinsloo, 2020).

El Grupo Independiente de Expertos de Alto Nivel sobre IA (2019) considera que la explicabilidad está directamente relacionada con los principios de transparencia, rendición de cuentas y respeto a los derechos fundamentales. En este marco, se especifica que los sistemas de IA deben ser no solo técnicamente comprensibles para expertos, sino también interpretables para los usuarios finales, incluidos los estudiantes, docentes y personal administrativo.

¹ Para distinguir entre transparencia y explicabilidad proponemos la distinción operativa del NIST (2023). La Transparencia responde “qué pasó” en el sistema (qué es, cómo se usa, qué información relevante existe y quién responde). La Explicabilidad responde “cómo” se tomó la decisión o se generó el resultado (mecanismos subyacentes, factores principales).

En el ámbito educativo, la explicabilidad es especialmente crítica debido al creciente uso de sistemas de analítica de aprendizaje (LA), sistemas de tutoría inteligente (ITS) y herramientas de evaluación automatizada. La falta de explicabilidad puede generar múltiples riesgos, desde la pérdida de confianza en los sistemas hasta la reproducción de sesgos ocultos en evaluaciones, recomendaciones o rutas de aprendizaje personalizadas. Los sistemas opacos dificultan la validación externa, limitan la posibilidad de identificar errores algorítmicos y obstaculizan el desarrollo de una cultura de pensamiento crítico. Asimismo, si las decisiones generadas por estos sistemas no son comprensibles, se corre el riesgo de que los estudiantes pierdan el control sobre sus procesos de aprendizaje o que los docentes deleguen excesivamente su criterio pedagógico en la máquina.

Para abordar estos desafíos, la IA explicable (XAI) ha emergido como un campo especializado de investigación que busca desarrollar modelos capaces de justificar sus resultados en términos comprensibles, sin sacrificar eficiencia ni precisión. Según Doshi-Vélez y Kim (2017), un sistema de IA explicable debe permitir a los usuarios identificar cuáles variables fueron consideradas, cómo se ponderaron y por qué se llegó a una conclusión determinada. De este modo, XAI juega un papel clave, toda vez que permite que los usuarios comprendan las bases de las decisiones algorítmicas y participen activamente en su propio aprendizaje (Pontificia Academia para la Vida, 2020). Si las universidades adoptan IA sin explicabilidad ni control humano, corren el riesgo de convertir la educación en un proceso mecánico que debilita la creatividad y la toma de decisiones. Esta preocupación es compartida tanto por investigadores en el campo de la educación y la tecnología como por organismos internacionales (UNESCO, 2019; 2021; OCDE, 2021) que han desarrollado y propuesto principios éticos para el uso responsable de la IA.

La falta de transparencia en los sistemas de IA puede generar una "caja negra" algorítmica que impide identificar errores, reproducir resultados y comprender si las decisiones automatizadas están afectando injustamente a determinados grupos. Este riesgo resulta particularmente preocupante en contextos de alta sensibilidad, como la evaluación del desempeño estudiantil, la predicción del riesgo académico o la asignación de recursos educativos.

Desde el plano educativo, la transparencia también cumple una función pedagógica. Cuando los estudiantes comprenden cómo funcionan los algoritmos desarrollan pensamiento crítico y conciencia digital; de igual manera, cuando los docentes conocen los procesos mediante los cuales operan estos sistemas, pueden integrarlos de manera más responsable en su práctica educativa sin renunciar a su juicio profesional.

Este tipo de explicabilidad no solo tiene un valor técnico, sino también pedagógico y ético. Como sostienen Nguyen et al. (2023), cuando los estudiantes entienden cómo funcionan los algoritmos, desarrollan mayores niveles de confianza, autonomía y capacidad crítica para interactuar con las tecnologías, en lugar de asumir sus resultados como verdades absolutas.

La UNESCO (2021) también insiste en que los sistemas de IA deben permitir la trazabilidad de sus decisiones, especialmente cuando estas afectan derechos, oportunidades o trayectorias educativas.

Asimismo, la UE identifica la transparencia como requisito de IA fiable —incluida la explicabilidad y la comunicación adecuada a los afectados—, mientras que el *NIST* exige trazabilidad y documentación para apoyar interpretaciones y auditoría (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; *NIST*, 2023). En el campo educativo, esto implica que tanto los estudiantes como los docentes deben contar con información clara sobre el tratamiento de datos, las finalidades y los criterios de decisión, así como con vías efectivas para intervenir, corregir o impugnar decisiones automatizadas.

La explicabilidad se relaciona con el derecho a la información, con el principio de transparencia y con las garantías de debido proceso cuando exista evaluación o perfilamiento automatizado.

Finalmente, la explicabilidad no es un lujo técnico, por el contrario, es una condición ética y epistemológica para la justicia educativa. En contextos de desigualdad, los sistemas opacos tienden a reproducir sesgos existentes, sin ofrecer herramientas para su detección o corrección. Solo una IA que pueda explicar sus procesos puede ser sometida a escrutinio ético, validación pedagógica y mejora democrática.

Como componente del principio de transparencia, la explicabilidad adquiere matices distintos según el tipo de sistema utilizado. En la GenAI resulta necesario comprender los criterios que influyen en la generación de respuestas o contenidos. En la IA Agéntica, además de explicar el contenido generado, debe poder justificarse la secuencia de decisiones, herramientas utilizadas y acciones ejecutadas para alcanzar un determinado objetivo.

2. 2. 4. Responsabilidad y Rendición de Cuentas

La responsabilidad y la rendición de cuentas constituyen principios indispensables para garantizar que el uso de la IA en la educación permanezca bajo supervisión humana y que existan mecanismos claros para responder por las consecuencias derivadas de su utilización. En este sentido, no basta con que los sistemas sean técnicamente eficaces; también es necesario definir quién responde por sus decisiones, cómo se supervisan sus resultados y qué mecanismos existen para corregir errores, reparar posibles daños y garantizar el respeto de los derechos de las personas.

En paralelo, la responsabilidad requiere que existan estructuras organizacionales y normativas capaces de supervisar el uso de la IA y responder por sus impactos. La UNESCO (2021) enfatiza que las instituciones educativas deben contar con políticas de gobernanza que incluyan principios éticos, figuras responsables, auditorías periódicas y mecanismos de apelación para los usuarios afectados por decisiones algorítmicas.

Nguyen et al. (2023) señalan que la responsabilidad también debe incluir un componente proactivo: no basta con reaccionar ante fallos o sesgos, sino que se deben prevenir daños mediante procesos de revisión constante, mejora iterativa y participación de múltiples actores en la toma de decisiones. En este sentido, la responsabilidad implica tanto una disposición institucional como una cultura ética compartida. Como señala Floridi et al. (2018), ningún sistema de IA puede sustituir la responsabilidad última de las personas ni su obligación de rendir cuentas por las decisiones adoptadas con apoyo de estas tecnologías.

La encíclica *Magnifica Humanitas* refuerza este planteamiento al señalar que el desarrollo y la implementación de la IA requieren estructuras claras de responsabilidad y supervisión, de modo que las decisiones tomadas con apoyo de estos sistemas permanezcan siempre bajo la autoridad y el discernimiento de las personas (León XIV, 2026).

Por último, la Pontificia Academia para la Vida (2020) advierte que la transparencia y la responsabilidad son pilares para construir una IA verdaderamente ética, entendida como aquella que respeta las leyes y promueve el florecimiento humano. En educación, esto significa garantizar que toda implementación tecnológica esté acompañada de procesos deliberativos, mecanismos de revisión y responsabilidad compartida entre desarrolladores, autoridades institucionales y usuarios.

La responsabilidad institucional debe abarcar tanto los resultados producidos por sistemas generativos como las acciones ejecutadas por sistemas agénticos. Mientras que en la GenAI la rendición de cuentas suele centrarse en los contenidos generados, en la IAA también debe extenderse a las decisiones automatizadas, a las acciones ejecutadas por el sistema y a sus consecuencias dentro de los procesos educativos y administrativos.

2. 2. 5. Gobernanza y Supervisión

La gobernanza de la IA en el ámbito educativo no se refiere únicamente a aspectos técnicos o administrativos, sino que constituye un componente ético y estratégico fundamental para asegurar que la tecnología esté alineada con los valores y objetivos de las instituciones educativas. Gobernar la IA significa establecer marcos normativos, estructuras organizacionales, procesos participativos y mecanismos de monitoreo que permitan anticipar, controlar y corregir los impactos sociales, pedagógicos y legales del uso de sistemas automatizados.

La UNESCO (2021) define la gobernanza de la IA como un conjunto de principios, regulaciones y prácticas que permiten orientar su desarrollo hacia el bienestar humano, considerando el respeto a los derechos fundamentales, la inclusión social, y la diversidad cultural. Este organismo subraya que la gobernanza debe ser dinámica, abierta al diálogo interdisciplinario, y sensible a los contextos locales, especialmente en sectores como la educación, donde las decisiones automatizadas pueden tener consecuencias duraderas.

Por su parte, la OCDE (2022) propone una visión de gobernanza responsable basada en la rendición de cuentas, la transparencia y la participación de las partes interesadas. Desde esta perspectiva, las instituciones educativas deben actuar activamente en su regulación interna, asegurando la supervisión ética de los procesos de diseño, implementación y evaluación de la IA. La Pontificia Academia para la Vida (PAV, 2020) refuerza esta visión al señalar que la supervisión humana significativa es una condición necesaria para preservar la dignidad, la libertad y la justicia en cualquier contexto donde opere la IA.

La gobernanza también implica definir responsabilidades institucionales claras. Esto incluye la designación de comités o consejos éticos especializados en IA, la incorporación de responsables institucionales de protección de datos con competencias en temas digitales, y la elaboración de lineamientos internos que regulen el uso de tecnologías emergentes.

Esta perspectiva encuentra un respaldo reciente en la encíclica *Magnifica Humanitas*, donde León XIV señala que la creciente complejidad de estas tecnologías hace indispensable fortalecer mecanismos de gobernanza capaces de garantizar la transparencia, la rendición de cuentas y la protección de la dignidad humana durante todo su ciclo de vida (León XIV, 2026).

Los marcos de gobernanza propuesto por la UE y del EUA, que explicaremos más adelante, sitúan la gobernanza como condición de legitimidad y control: demandan roles claros, rendición de cuentas, participación de actores relevantes, documentación y revisión continua de decisiones a lo largo del ciclo de vida (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023; ISO 42001; ISO 23894). Este enfoque plantea que la gobernanza no debe concebirse como una política vertical o un manual exclusivamente técnico, sino como un proceso vivo e iterativo que equilibre innovación, seguridad y justicia. Otros marcos de gobernanza muy importantes en las empresas son el ISO 42001 y el ISO 23894. Estos documentos técnicos, junto con las antes mencionadas NIST y EU Act, son las más adoptadas para implementar, mantener y mejorar un Sistema de Gestión de la IA (SGIA) en las organizaciones.

A nivel práctico, la gobernanza se traduce en acciones como la implementación de auditorías de sesgo algorítmico, el monitoreo del cumplimiento de principios como la equidad o la explicabilidad, y la creación de espacios de diálogo con docentes, estudiantes y especialistas externos. La gobernanza también se articula con la alfabetización digital crítica, ya que sin comprensión de cómo operan los sistemas, es imposible ejercer una supervisión efectiva o formular una crítica informada.

Finalmente, la gobernanza de la IA en educación es un compromiso con la responsabilidad institucional y el desarrollo ético de la comunidad universitaria. Implica reconocer que la implementación de tecnologías debe estar guiada por principios democráticos, derechos

humanos y una visión pedagógica centrada en las personas, no en la eficiencia operativa. Como argumenta el marco AI4People, una gobernanza efectiva requiere tanto la infraestructura jurídica como una infraestructura ética complementaria, capaz de responder a los riesgos dinámicos del entorno digital que la norma no siempre alcanza a tiempo (Floridi et al., 2018).

La necesidad de gobernanza se vuelve especialmente relevante en los sistemas de IA agéntica debido a su capacidad para planificar y ejecutar acciones con distintos grados de autonomía. Mientras la GenAI requiere controles sobre la calidad y fiabilidad de los contenidos generados, la IAA exige mecanismos adicionales de autorización, monitoreo, trazabilidad y supervisión humana continua.

2. 2. 6. Inclusión y Equidad

Un principio fundamental en la ética de la IA es que su desarrollo debe garantizar que todas las personas puedan beneficiarse de sus avances sin generar desigualdades. En este sentido, el *Rome Call for AI Ethics* enfatiza que la educación es un derecho humano fundamental y que las tecnologías emergentes deben promover la equidad en su acceso y uso (Pontificia Academia para la Vida, PAV 2020).

Asimismo, este llamado global resalta la necesidad de que la IA incluya a las comunidades marginadas, garantizando que la automatización y la personalización del aprendizaje no generen brechas educativas. Para ello, los diseñadores de sistemas educativos deben asegurarse de que sus plataformas sean accesibles para todos los estudiantes, incluidas personas con discapacidad o con acceso limitado a infraestructura tecnológica. De lo contrario, se corre el riesgo de que la IA refuerce desigualdades estructurales preexistentes en los sistemas educativos (Pontificia Academia para la Vida, PAV 2020).

Además, el *Rome Call for AI Ethics* (Pontificia Academia para la Vida, PAV 2020) insiste en que la IA debe diseñarse y usarse de modo que beneficie a toda la humanidad, en particular a los más desfavorecidos. Esto implica priorizar proyectos que atiendan brechas estructurales, como el abandono escolar, la discriminación interseccional o la marginación digital.

La inclusión también se vincula con la participación: las comunidades educativas deben tener voz en el diseño y evaluación de los sistemas, de modo que las soluciones tecnológicas no se impongan de manera vertical, sino que respondan a las necesidades reales de sus usuarios. Solo a través de una IA sensible a la justicia social es posible avanzar hacia un modelo educativo verdaderamente humanista e inclusivo.

Además, se proponen que estas tecnologías eviten reproducir desigualdades existentes y contribuyan activamente a su superación. La equidad no puede ser entendida simplemente

como acceso a dispositivos o conectividad, sino como la garantía de que todos los estudiantes, sin importar su origen socioeconómico, género, etnicidad, discapacidad o localización geográfica, puedan beneficiarse de la IA de manera justa, significativa y contextualizada.

La UNESCO (2021) sostiene que una IA ética debe promover la equidad como un valor transversal, asegurando que los beneficios de estas tecnologías no estén concentrados en ciertos sectores de la población ni generen nuevas formas de exclusión. En el ámbito educativo, esto implica repensar los algoritmos, los conjuntos de datos y las decisiones automatizadas desde una perspectiva de justicia distributiva y reconocimiento de la diversidad.

En particular, Nguyen et al. (2023) argumentan que el desarrollo de sistemas de IA inclusivos requiere mejoras técnicas y una revisión crítica de los supuestos pedagógicos que los sustentan. Por ejemplo, deben evitarse modelos que asumen que todos los estudiantes aprenden de la misma manera, o que privilegian indicadores cuantitativos de éxito sin considerar factores emocionales, culturales o comunitarios.

Autores como Eubanks (2018), han advertido que muchos sistemas automatizados, aunque diseñados con criterios de eficiencia, terminan amplificando las desigualdades estructurales, especialmente cuando se aplican sin considerar los contextos culturales o históricos de los usuarios. En educación, esto se refleja en sistemas de recomendación que reproducen sesgos de género o raza, en plataformas adaptativas que reducen el contenido para estudiantes en desventaja, o en procesos de selección algorítmica que penalizan trayectorias no lineales.

La implementación de sistemas de IAED puede incorporar sesgos en los datos de entrenamiento o en los algoritmos, lo que podría derivar en discriminación contra determinados grupos de estudiantes, por ejemplo, por origen étnico o condición socioeconómica (Cordón, 2023). Por ejemplo, si se toman decisiones de admisión basadas en IAED, la existencia de sesgos puede traducirse en una reducción injustificada de becas u oportunidades para ciertos grupos de estudiantes (Liu et al., 2023).

Asimismo, otro ejemplo de sesgo puede observarse en los ITS y en otros modelos de personalización del aprendizaje, que pueden verse influidos por sesgos implícitos en los datos de los que se nutren, generando resultados desiguales para estudiantes de distintos orígenes (Prinsloo, 2020). Por lo anterior, resulta necesario crear mecanismos y políticas institucionales para asegurar que las decisiones apoyadas por algoritmos en contextos educativos no discriminen a determinados grupos (Alonso-Rodríguez, 2024).

La equidad también exige accesibilidad. Según la OCDE (2022), los sistemas de IA deben diseñarse desde un enfoque de diseño universal, lo que implica que puedan ser utilizados por personas con distintas capacidades físicas, cognitivas y lingüísticas. Esto incluye desde la

adaptación de interfaces, hasta la generación de contenidos en lenguas indígenas o la integración de tecnologías de asistencia.

Asimismo, el requisito de diversidad, no discriminación y equidad de la UE, junto con la gestión explícita de riesgos de sesgo y daño del NIST, exigen evaluar continuamente cómo los sistemas afectan a poblaciones vulnerables e incorporar mecanismos de participación y retroalimentación para detectar sesgos, errores o efectos colaterales no previstos (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023).

Este enfoque sugiere que la equidad no es un resultado automático del desarrollo tecnológico, sino una construcción institucional que requiere voluntad política, sensibilidad social y compromiso pedagógico.

Los sesgos presentes en sistemas generativos pueden reproducirse en textos, recomendaciones o contenidos educativos. En sistemas agénticos, dichos sesgos pueden amplificarse cuando las recomendaciones son traducidas en acciones o decisiones automatizadas que afectan directamente a estudiantes y docentes.

2. 2. 7. Sostenibilidad y Proporcionalidad

El principio de sostenibilidad y proporcionalidad en el uso de la IA en la educación superior plantea la necesidad de equilibrar el desarrollo tecnológico con el respeto al entorno ecológico, la racionalidad en el uso de recursos y la adecuación contextual de las soluciones implementadas. No toda innovación tecnológica es éticamente deseable si conlleva un costo ambiental elevado o una implementación desproporcionada en relación con sus beneficios reales.

La UNESCO (2021) establece que la IA debe contribuir al desarrollo sostenible, de acuerdo con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030, promoviendo prácticas responsables en su diseño, uso y ciclo de vida. Esto implica considerar el impacto energético de los sistemas, los residuos digitales generados y la huella ecológica de las infraestructuras tecnológicas, especialmente en instituciones educativas que adoptan plataformas a gran escala.

Desde esta perspectiva, la sostenibilidad también abarca la justicia intergeneracional: las decisiones tecnológicas actuales deben contemplar su impacto en las generaciones futuras, tanto en términos medioambientales como educativos. Como señalan Nguyen et al. (2023), la expansión acrítica de sistemas de IA en educación puede acarrear costos no visibles a corto plazo, como el consumo energético intensivo de servidores, la obsolescencia tecnológica acelerada y la dependencia estructural de proveedores privados que condicionan la autonomía institucional.

Más allá del consumo de energía, la sostenibilidad de la IA también está relacionada con otras dimensiones humanas, sociales, culturales, económicas y ambientales (Martínez, 2024). El principio de proporcionalidad, por su parte, exige que la solución tecnológica sea adecuada, necesaria y no excesiva frente al problema que busca resolver. Inspirado en principios jurídicos de razonabilidad y balance, este criterio busca evitar la sobre-regulación algorítmica, el uso de sistemas invasivos para tareas que podrían resolverse con intervenciones humanas simples, o la implementación de tecnologías cuya complejidad supera las capacidades de los usuarios.

La OCDE (2022) y el Grupo de Expertos de Alto Nivel sobre IA de la Comisión Europea (2019) coinciden en que la IA debe aplicarse con un enfoque de proporcionalidad digital que valore la pertinencia de cada sistema según el contexto educativo específico, la madurez tecnológica de la institución y las necesidades reales de los usuarios. La automatización no debe convertirse en un fin en sí mismo, sino en un medio que agregue valor sin desplazar innecesariamente el juicio pedagógico o aumentar la carga tecnológica.

Por su parte, la proporcionalidad y la sostenibilidad pueden entenderse como exigencias de gobernanza responsable: la UE incorpora el bienestar social y ambiental como requisito de IA fiable, y el *NIST* recomienda gestionar riesgos considerando costos, beneficios y efectos a lo largo del tiempo (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; *NIST*, 2023). Esto implica preguntarse si el sistema es necesario y proporcional para el fin educativo, si es replicable y mantenible, si su financiamiento es transparente y si responde a una necesidad genuina y prioritaria de la comunidad universitaria.

Por último, el *Rome Call for AI Ethics* (Pontificia Academia para la Vida, PAV 2020) advierte que la tecnología debe estar al servicio de una ecología integral, respetando el entorno natural, los ritmos humanos y la justicia social. En este sentido, la sostenibilidad y la proporcionalidad también son principios de contención ética: nos recuerdan que no todo lo posible es deseable, y que una educación verdaderamente humanista requiere tecnologías sobrias, conscientes y responsables.

2. 2. 8. No Maleficencia y Prevención de Daños

El principio de no maleficencia, originado en la ética médica, establece que toda intervención debe evitar causar daño. La IA no puede ser éticamente válida si compromete la integridad, dignidad o desarrollo de las personas involucradas. El Grupo de Expertos de Alto Nivel sobre IA (2019) y la OCDE (2022) coinciden en que esta precaución debe guiar especialmente el uso de IA en sectores de alto impacto como la educación, donde las decisiones afectan trayectorias de vida. En este sentido, la UNESCO (2021) señala que una IA ética debe identificar, prevenir y mitigar los riesgos potenciales que puedan derivarse de su aplicación, tanto a nivel físico, psicológico como social. También se debe considerar la prevención de daños incluida afectación reputacional o psicológica por vigilancia, o decisiones erróneas.

Esto incluye desde sesgos algorítmicos que perpetúan desigualdades, hasta impactos negativos en la salud mental, vigilancia excesiva o deshumanización de los procesos pedagógicos.

Nguyen et al. (2023) sostienen que muchos de los daños asociados a la IA educativa no son evidentes de forma inmediata, sino que emergen con el tiempo, especialmente en contextos donde los sistemas son percibidos como injustos, invasivos o incomprensibles. Estos efectos pueden manifestarse como disminución de la motivación, ansiedad académica, desconfianza institucional o exclusión de estudiantes con trayectorias atípicas.

Finalmente, tanto el enfoque de IA fiable de la UE como el *AI RMF* del *NIST* integran la prevención del daño como criterio transversal: exigen robustez y seguridad técnica, así como evaluación de impactos y monitoreo continuo de efectos no deseados a lo largo del ciclo de vida (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; *NIST*, 2023). Para ello, pueden emplearse pruebas y simulaciones, análisis de impacto, revisiones externas y mecanismos de seguimiento que permitan detectar desviaciones, corregirlas y prevenir afectaciones a personas y colectivos.

A nivel operativo, la prevención de daños requiere crear espacios de deliberación anticipada, formar a los equipos de desarrollo y al personal académico en ética tecnológica, y establecer protocolos que limiten el uso de IA en contextos de alta vulnerabilidad o sin supervisión humana. Además, es necesario diseñar tecnologías que permitan pausas, desconexión y respeto al ritmo humano, reconociendo que el exceso de automatización puede tener consecuencias perjudiciales incluso si es técnicamente eficiente.

El principio de no maleficencia no es una barrera al desarrollo, es un llamado a que la innovación educativa se construya desde el cuidado, la empatía y el compromiso ético.

Ampliando la mira de este principio, el *Rome Call for AI Ethics* (Pontificia Academia para la Vida, 2020) resalta que evitar el daño no es suficiente: también se debe promover activamente el bien. Desde este punto de vista, prevenir el daño es, en última instancia, una expresión del respeto por la vida y la dignidad de las personas, y una condición indispensable para que la IA tenga un lugar legítimo en el horizonte pedagógico.

Actualmente, estos riesgos varían según la tecnología implementada. En el caso de la GenAI, los daños potenciales suelen manifestarse mediante desinformación, alucinaciones, sesgos o afectaciones a la integridad académica. Por su parte, la IAA añade riesgos como la ejecución de acciones inadecuadas, errores en procesos automatizados o decisiones que impacten negativamente en estudiantes sin una supervisión humana suficiente.

2. 2. 9. Supervisión Humana

Este principio establece que toda decisión relevante tomada por o con apoyo de un sistema automatizado de IA debe estar sujeta a una revisión, interpretación o validación humana, especialmente cuando afecta derechos, trayectorias o experiencias formativas de estudiantes y docentes. La UNESCO (2021) plantea que la supervisión humana debe ir más allá de ser simbólica o superficial, debiendo ser efectiva y responsable. Esto implica que las personas involucradas tengan no solo el derecho, sino también la capacidad, la formación y la autoridad para intervenir, modificar o revertir decisiones algorítmicas.

El Grupo de Expertos de Alto Nivel sobre IA (2019), a través de sus directrices para una IA confiable, introduce el concepto de *human agency* como componente esencial. Este concepto subraya que las tecnologías no deben sustituir la capacidad de acción humana, sino fortalecerla. En educación, esto implica que los sistemas deben ser comprensibles, auditables y adaptables a juicio docente. Además, cualquier proceso de toma de decisiones automatizado debe ofrecer explicaciones claras y permitir cuestionamientos por parte de los usuarios afectados. Nguyen et al. (2023) destacan que la confianza de los estudiantes en los sistemas de IA depende en gran medida de su percepción de que existe supervisión humana efectiva. Cuando los estudiantes sienten que las decisiones están completamente delegadas a máquinas, sin posibilidad de apelación o revisión, su sentido de agencia disminuye, pudiendo generar frustración, desmotivación o resistencia al aprendizaje.

En consecuencia, la supervisión significativa requiere más que una presencia humana formal: la UE exige supervisión y agencia humanas, y el NIST recomienda estructurar roles, capacidades y controles para evitar automatización acrítica (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023). Esto supone mecanismos institucionales permanentes —comités de revisión, canales de retroalimentación, evaluaciones participativas de impacto y formación—, de modo que quienes supervisan comprendan el sistema y cuenten con potestad real para detenerlo, ajustarlo o impugnar sus resultados.

La OCDE (2022) también resalta que la supervisión significativa requiere una gobernanza institucional clara: quién supervisa, con qué criterios, con qué frecuencia y con qué consecuencias. La transparencia organizacional es clave para garantizar que el principio sea simplemente efectivamente practicado. Esta perspectiva coincide con la propuesta de León XIV, quien sostiene que corresponde siempre a la inteligencia humana, acompañada de la conciencia y la libertad, conducir las innovaciones tecnológicas y establecer responsablemente sus límites.

En suma, la supervisión humana significativa no se limita a vigilar a las máquinas, sino que implica preservar el sentido ético, pedagógico y contextual de las decisiones educativas. Su finalidad es prevenir errores técnicos, proteger la dimensión humana del aprendizaje, el

diálogo entre saberes, y el acompañamiento formativo que ninguna tecnología puede reemplazar.

La supervisión humana constituye una exigencia transversal para ambos tipos de sistemas. Sin embargo, adquiere una relevancia aún mayor en la IA agéntica, debido a que estos sistemas no sólo generan información, sino que pueden intervenir activamente en procesos institucionales mediante la ejecución de acciones orientadas al cumplimiento de objetivos específicos.

2. 2. 10. Alfabetización en IA

La alfabetización en IA se ha convertido en una condición ética esencial para el desarrollo y uso responsable de tecnologías automatizadas en la educación superior. Este principio sostiene que, para que los sistemas de IA sean utilizados de forma crítica, informada y justa, es necesario que todos los actores educativos —estudiantes, docentes, personal administrativo y directivo— desarrollen competencias básicas para comprender su funcionamiento, implicaciones y límites.

Es oportuno incluir una dimensión formativa con base en el Triángulo de la Irreducibilidad Humana: prácticas para reconocer y cuidar la relación con otros (empatía), para deliberar sobre dilemas y responsabilidades en el uso de sistemas automatizados (juicio ético), y para generar alternativas y justificar decisiones con sentido (pensamiento divergente). Sin esta formación, la interacción con IA tiende a volverse acrítica y dependiente.

Organismos como la UNESCO (2021) y la OCDE (2022) subrayan que, sin una comprensión básica de cómo se entrenan los modelos, qué sesgos arrastran y qué límites poseen, la comunidad universitaria corre el riesgo de aceptar la IA como caja negra, reforzando dependencias y reproduciendo desigualdades.

En el contexto educativo, la falta de alfabetización algorítmica puede reforzar relaciones de dependencia y verticalidad, donde los sistemas se aceptan de forma acrítica o se utilizan de manera inapropiada. Asimismo, los marcos de IA fiable y de gestión de riesgos advierten que asumir neutralidad tecnológica invisibiliza sesgos y decisiones de diseño; por ello, la comprensión, documentación y discusión pública de los sistemas es condición de uso responsable (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023).

Nguyen et al. (2023) han demostrado que el nivel de comprensión que tienen los estudiantes sobre cómo funciona la IA educativa afecta directamente su confianza, su percepción de justicia y su motivación. Esta relación también se observa en los docentes: aquellos que han sido formados en los fundamentos de IA tienen más herramientas para adaptar sus prácticas pedagógicas, cuestionar las recomendaciones de los sistemas y acompañar a sus alumnos de manera crítica.

Desde una perspectiva institucional, el *Rome Call for AI Ethics* (Pontificia Academia para la Vida, PAV 2020) afirma que una IA verdaderamente humanista sólo puede desarrollarse cuando las comunidades participan activamente en su diseño y gobernanza. Esto exige construir capacidades colectivas para comprender y dialogar sobre los sistemas tecnológicos, y no relegar su gestión a especialistas externos.

Por ello, la alfabetización en IA debe integrarse en los planes de estudio de educación superior, no sólo como parte de programas especializados en tecnología, sino como componente transversal de la formación ética y ciudadana. Asimismo, deben generarse políticas de desarrollo profesional para docentes y personal institucional, que les permitan evaluar críticamente los sistemas de IA y participar en su adaptación y mejora continua.

Esta alfabetización va más allá del mero manejo de herramientas: implica cuestionar la lógica de los algoritmos, analizar su impacto social y ejercer una supervisión responsable. El marco *Critical AI Literacy* de la Open University (2025) ilustra bien este enfoque, al articular niveles progresivos de competencia —desde la familiarización con conceptos básicos hasta la aplicación experta y la reflexión ética— y al vincular cada nivel con resultados de aprendizaje concretos y evaluables.

Existen, además, experiencias institucionales que traducen estos principios en protocolos operativos. Un ejemplo es la Plantilla de Protocolo para el Uso de la IA en Actividades Formativas adoptada por la Universitat Abat Oliba CEU (2023), donde los estudiantes documentan qué herramientas emplean, cómo mitigan sesgos y qué reflexiones éticas extraen del proceso. Este formato —que exige describir funciones, justificar decisiones y comprobar resultados frente a métodos tradicionales— convierte la alfabetización en un ejercicio vivencial de responsabilidad académica.

La alfabetización crítica cierra el círculo de los principios porque convierte la ética en acción cotidiana y la hace operativa; sin ella, los demás postulados quedarían en simples declaraciones de intención. Formar usuarios competentes plantea un reto sistémico que rebasa el aula y exige revisar la arquitectura institucional en su conjunto. Aquí entra en juego la perspectiva macro de la gobernanza: es decir, cómo articular normas, procesos y actores para que los principios éticos de uso de IA en el plano pedagógico se reflejen en políticas, infraestructuras y modelos de gobernanza.

Para lograr hacer vida estos principios éticos dentro de las IES se proponen y se recogen propuestas de organismos nacionales e internacionales. En primer lugar, *The Office of Educational Technology Department of Education* de EE. UU. establece la premisa de poner siempre en el centro al educador; a esto queremos añadir la sugerencia de la UE sobre marcos verificables de evaluación ética y gestión de riesgos: Directrices Éticas para una IA Fiable UE (2019) y el *AI Risk Management Framework* (NIST, 2023). La alfabetización en IA debe

permitir que docentes y estudiantes comprendan las diferencias entre sistemas generativos y sistemas agénticos. Esta comprensión resulta indispensable para identificar tanto las oportunidades como los riesgos asociados a la generación automatizada de contenidos y a la creciente autonomía operativa de algunos sistemas de IA.

La alfabetización en IA debe permitir que docentes y estudiantes comprendan las diferencias entre sistemas generativos y sistemas agénticos. Esta comprensión resulta indispensable para identificar tanto las oportunidades como los riesgos asociados a la generación automatizada de contenidos y a la creciente autonomía operativa de algunos sistemas de IA.

2. 3. Premisa Fundamental: la Centralidad del Docente

La propuesta ACE (*Always Center Educators*) se contempla en el diagnóstico que emitió la Oficina de Tecnología Educativa del Departamento de Educación de los Estados Unidos (2023). En dicho informe se presenta la premisa de que, para tener éxito con la IA como un complemento del aprendizaje y la enseñanza, se debe siempre centrar la atención en los educadores. En términos prácticos, “aplicar el ACE en IA significa mantener una visión humanista de la enseñanza como prioridad” (U.S. Department of Education, Office of Educational Technology, 2023, p. 29).

Para definir con mayor claridad el papel central de los educadores (ACE), retomamos la propuesta de una IA con control humano y planteamos una pregunta: ¿en qué punto del ciclo educativo deben ubicarse los docentes como actores clave?

El informe integra tres momentos del ciclo educativo donde el docente debe estar en el centro, y no ser sustituido por la IA: (1) la enseñanza directa (*Doing Teaching*), (2) la preparación y reflexión sobre la enseñanza (*Preparing for and Reflecting on Teaching*), y (3) el diseño, selección y evaluación de herramientas para la enseñanza y el aprendizaje (*Designing, Selecting, and Evaluating Tools*).

Aunque el informe no detalla una metodología operativa exhaustiva, su estructura implícita sugiere una lógica circular de mejora continua, en la que el juicio pedagógico no sólo inicia, sino regula y corrige los ciclos de adopción tecnológica en educación (U. S. *Department of Education*, Office of Educational Technology, 2023).

Figura 5. Propuesta ACE para el ciclo educativo.



Nota: Elaboración de U. S. Department of Education, Office of Educational Technology (2023).

Esta disposición gráfica enfatiza que, en cualquier punto del ciclo educativo, los educadores deben permanecer al centro del diseño, implementación y evaluación de soluciones con IA. Dicho en otras palabras, la actividad docente en todas sus fases se articula con una premisa central: el docente debe permanecer siempre en el centro, como agente activo y deliberativo, que puede servirse de la IA como una herramienta.

La inclusión del profesorado en todas las fases del ciclo educativo—diseño, implementación y ajuste— no solo refuerza su agencia pedagógica, sino que constituye una salvaguarda crítica contra los sesgos algorítmicos y la despersonalización del proceso educativo (Doshi-Vélez & Kim, 2017; Gunning & Aha, 2019).

El principio de ACE establece como eje pedagógico la iteración con docentes, la formación continua y la amplificación del juicio pedagógico. De este modo, la centralidad del docente es una estrategia concreta para garantizar una implementación ética de la IA en las aulas (U. S. Department of Education, Office of Educational Technology, 2023, p. 29).

A partir de una lectura crítica del informe de la Oficina de Tecnología Educativa del Departamento de Educación de los Estados Unidos y en diálogo con la literatura académica reciente (Chan, 2023; Holmes et al., 2022; Williamson & Eynon, 2020), se proponen tres ideas: primero, que la IA amplifique y no sustituya el juicio docente; segundo, que los sistemas se co-diseñen con el profesorado desde las etapas iniciales; y tercero, que exista formación continua en el uso ético, pedagógico y técnico de esta tecnología. Esta última

premisa hace referencia al principio de alfabetización crítica en IA. Estas tres ideas no aparecen como una lista cerrada en el documento original del Departamento de Educación, al contrario, nos permiten pensar la IA no como reemplazo, sino como extensión del proceso de enseñanza-aprendizaje (Holmes et al., 2022; Williamson & Eynon, 2020; Chan, 2023).

Holmes et al. (2021) abogan por una integración que preserve el juicio profesional del docente sobre las decisiones automatizadas de la IA, resaltando la necesidad de intervenciones humanas reflexivas. Por su parte, Williamson y Eynon (2020) cuestionan la supuesta neutralidad de los sistemas algorítmicos y llaman a una participación crítica del profesorado en su diseño e implementación, al advertir que todo sistema educativo mediado por IA refleja una determinada política del conocimiento. Finalmente, Chan (2023) propone modelos participativos de co-diseño donde el profesorado no solo es usuario final, sino coproductor de la tecnología, asegurando así su pertinencia, usabilidad y valor pedagógico real. En conjunto, estas contribuciones refuerzan la noción de que la IA educativa debe amplificar, no sustituir, la agencia docente, y que su incorporación ética exige mantener a las y los educadores en el centro del diseño y la toma de decisiones.

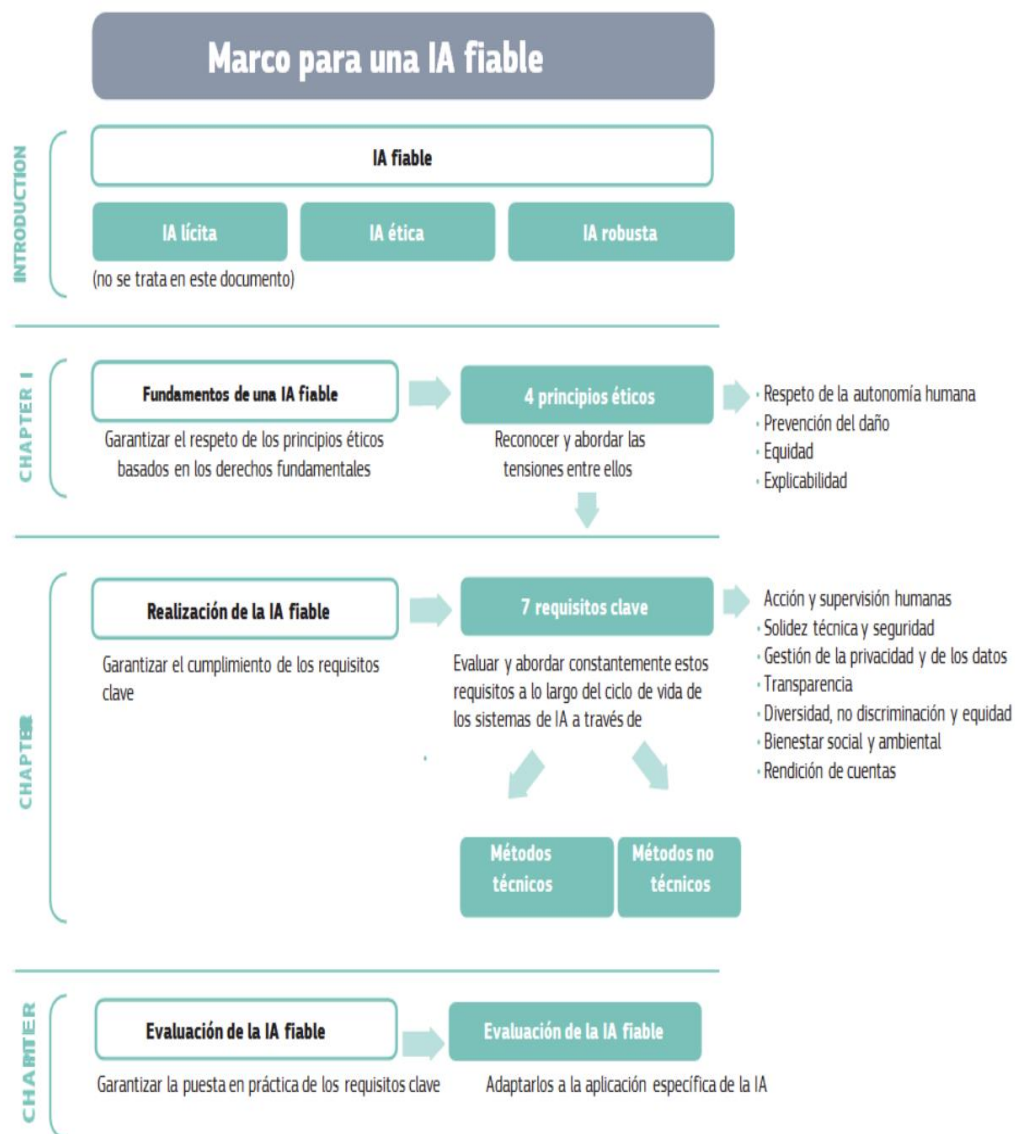
La propuesta ACE se alinea con el concepto de *Human-in-the-Loop* (Gunning & Aha, 2019), proponiendo que toda tecnología educativa integre mecanismos que garanticen supervisión docente significativa en cada etapa de su implementación y uso. Además, la propuesta ACE contribuye a la calidad educativa, la libertad académica y la responsabilidad del docente.

2. 4. Marcos de Evaluación Ética y Gestión de Riesgos

Las *Directrices para una IA fiable* de la UE establecen que la confianza es un requisito previo para que las sociedades y las economías adopten esta tecnología, maximicen sus beneficios y minimicen sus riesgos. Más adelante, el documento define la IA fiable como un estándar que debe sostenerse durante todo el ciclo de vida del sistema y que integra tres componentes necesarios: (a) una IA lícita, que asegure el cumplimiento normativo; (b) una IA ética, que garantice el respeto de principios y valores; y (c) una IA robusta, tanto en el aspecto técnico como en el social, para evitar daños accidentales incluso ante buenas intenciones (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019).

El enfoque de IA fiable de la Unión Europea propone cuatro principios éticos (respeto de la autonomía humana, prevención del daño, equidad y explicabilidad) y siete requisitos para orientar el diseño, desarrollo y despliegue de sistemas confiables, lo que permite traducir dichos principios en listas de verificación y mecanismos institucionales de seguimiento (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; ver Figura 6). El documento también hace referencia a las múltiples partes que intervienen en el proceso (desarrolladores, responsables del despliegue, usuarios finales y sociedad) e indica que su implementación puede apoyarse en métodos técnicos y no técnicos.

Figura 6. Marco para una IA fiable.



Nota: Tomada de *Ethics Guidelines for Trustworthy AI*, por la Comisión Europea, 2019 (<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>).

Por lo que se refiere al *AI Risk Management Framework* (Marco de gestión de riesgos) del NIST organiza la gestión de riesgos de IA como un ciclo continuo —gobernar, mapear, medir y manejar— que ayuda a identificar contextos de uso, medir riesgos, implementar controles y revisar decisiones conforme cambian los datos, los modelos y el entorno regulatorio (NIST, 2023).

En el caso de las IES, si adoptamos este marco podemos estructurar la gobernanza ética de la IA a partir de lo siguiente: delimitar finalidades, actores y contexto de uso (*map*); establecer roles, políticas y rendición de cuentas (*govern*); medir (*measure*) riesgos —

privacidad, sesgo, seguridad y transparencia— con evidencia; y gestionar (*manage*) controles con mejora continua (NIST, 2023).

Figura 7. Organize AI risk management activities at their highest level to govern, map, measure, and manage AI risk.



Nota: Tomada de Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1), por el National Institute of Standards and Technology, 2023 (<https://doi.org/10.6028/NIST.AI.100-1>). En el original: "Organize AI risk management activities at their highest level to govern, map, measure, and manage AI risk".

Ambos marcos —Directrices Éticas para una IA Fiable UE (2019) y el *AI Risk Management Framework del NIST* (2023)— convergen en una premisa metodológica: la deliberación ética debe tener un flujo continuo que permita repetir, materializarse en herramientas y puntos de control, de modo que la gobernanza sea verificable y no sólo declarativa (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023).

Estos marcos se relacionan con el principio de *accountability* y la gestión debida del riesgo en tratamiento de datos y automatización.

En el capítulo cuarto ahondaremos en la aplicabilidad de estos dos marcos a las IES para proteger la privacidad del estudiante y, concretamente la figura del consentimiento informado.

2. 5. Un Modelo Ético Iterativo que Pone al Docente en el Centro

Los principios éticos analizados en los apartados anteriores requieren algo más que declaraciones normativas para tener efectos reales en la práctica educativa. Para que estos principios orienten efectivamente el uso de la IA en la educación superior, es necesario que se integren en modelos de actuación que permitan su aplicación continua, contextualizada y pedagógicamente medida.

Por ello, la ética de la IA en la educación no puede entenderse como un marco estático, sino como un proceso dinámico que se construye y revisa en el tiempo, en función de los contextos de uso, de los riesgos emergentes y de las experiencias concretas de estudiantes y docentes. Esta visión coincide con lo señalado por la UNESCO (2021) y la OCDE (2021), que subrayan la necesidad de mecanismos de supervisión humana, rendición de cuentas y revisión permanente para garantizar un uso responsable de los sistemas de IA.

Dentro de este proceso ético dinámico, el docente ocupa un lugar central. Como se desarrolló en el apartado 2.3 el profesor no es un mero usuario de herramientas tecnológicas, sino el principal mediador entre la tecnología, el estudiante y el proceso formativo. El Departamento de Educación de los Estados Unidos (U. S. *Department of Education*, Office of Educational Technology, 2023), a través de la propuesta ACE (*Always Center Educators*), enfatiza que los docentes deben permanecer en el centro de las decisiones pedagógicas y tecnológicas, asegurando que la IA complemente, y no sustituya el juicio educativo humano.

Poner al docente en el centro implica reconocer su papel en tres momentos clave del proceso educativo: el diseño de las actividades de aprendizaje, la implementación y acompañamiento del proceso formativo, y la evaluación de los resultados. En cada una de estas etapas, el profesorado es quien puede identificar los efectos reales del uso de la IA, detectar posibles sesgos, explicar las decisiones algorítmicas y ajustar las prácticas pedagógicas en función de las necesidades de los estudiantes. De este modo, la supervisión humana deja de ser un principio abstracto y se convierte en una práctica educativa concreta.

Este enfoque refuerza la relación entre ética y alfabetización en IA. Cuando el docente actúa como mediador crítico, la ética se integra en el proceso formativo y se transmite al estudiantado no solo como un conjunto de normas, sino como una forma de comprender y cuestionar el uso de la tecnología. Tal como señala la UNESCO (2021), la alfabetización en IA es una condición *indispensable* para que las personas puedan ejercer su autonomía, comprender los sistemas que utilizan y participar activamente en su supervisión.

Consideramos que un modelo centrado en el docente permite articular los principios éticos de la IA con la práctica educativa cotidiana. Al situar al profesorado como eje del proceso, se garantiza que la ética no quede relegada a documentos institucionales o marcos regulatorios, sino que se viva y se revise de manera continua en el aula. Esta aproximación

sienta las bases para comprender la confianza como un valor transversal, clave para la gobernanza ética de la IA.

Al poner en diálogo la premisa ACE (2023) con los requisitos de IA fiable de la UE (2019) y el ciclo de gestión de riesgos del NIST (2022), es *posible* traducir un marco general de ética aplicada a una arquitectura operativa de consentimiento informado para las IES.

Por una parte, la propuesta del ACE aporta la centralidad pedagógica y la agencia docente; por otra, la UE y el *NIST* aportan criterios verificables y controles institucionales para prevenir automatización indebida, sesgos y opacidad (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; *NIST*, 2023). Estos modelos operan en todo el ciclo y se aplica de modo iterativo.

La premisa ACE aporta el “quién” —el profesorado como agente y no la IA—, mientras que los marcos de la UE y del *NIST* aportan el “cómo” operativo: requisitos y funciones de gobernanza que permiten documentar decisiones, evaluar impactos y corregir riesgos de forma iterativa. Por consiguiente, su convergencia demuestra que la aplicación de los principios éticos de la IA no es un ejercicio abstracto, sino un proceso institucional verificable que sostiene la confianza y habilita el consentimiento informado.

Ambas propuestas, puestas en conjunto, dan a las IES un sistema de mejora continua de las garantías y previenen daños ante los cambios tecnológicos.

2. 6. Confianza como Valor Transversal

Dos de las propuestas anteriores destacan el papel primordial que tiene la confianza en la correcta adopción de IA (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; U. S. *Department of Education*, 2023).

La confianza constituye un elemento esencial y transversal para el desarrollo ético de la IA en el ámbito educativo. Tanto propuestas institucionales como marcos internacionales coinciden en que los principios éticos: centrarse en el ser humano, proteger la privacidad, asegurar la supervisión humana y promover la equidad dependen de una base de confianza que los haga posibles (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; U. S. *Department of Education*, 2023). Consideramos que, sin confianza, cualquier marco ético queda incompleto y carece de legitimidad ante la comunidad universitaria.

Covey (2006) sostiene que la confianza es un pilar imprescindible para la cooperación a largo plazo, y la UNESCO (2021) la reconoce como condición para la adopción legítima y sostenible de los sistemas de IA. En el ámbito educativo, esto se traduce en que sin confianza sería inviable construir estructuras estables de rendición de cuentas y coordinación

interinstitucional, ya que los grupos involucrados difícilmente aceptarían decisiones algorítmicas percibidas como opacas o injustas.

Además, la confianza actúa como un catalizador que incrementa la disposición a compartir datos y facilita la participación de estudiantes, docentes y personal técnico en procesos de co-diseño de los sistemas educativos mediados por IA. La confianza se construye —no se decreta— al priorizar los derechos y necesidades de estudiantes, docentes y personal administrativo en cada decisión tecnológica (Chan, 2023).

Rosales Gallego (2024) refuerza esta visión al señalar que la confianza es condición necesaria para generar legitimidad y participación social. En su texto, indica que la confianza se ha convertido en principio rector de la regulación europea emergente, precisamente para garantizar la protección de derechos fundamentales y la alineación con valores humanos.

La confianza resulta especialmente pertinente en entornos universitarios, donde la gobernanza de la IA es técnica, ética, inclusiva y socialmente aceptada. En este contexto la confianza actúa como el “pegamento social” que permite el co-diseño, implementación y aceptación de la IA en entornos educativos.

Pasetti y de Oliveira (2024) aportan un enfoque complementario al advertir que la confianza debe entenderse en sentido propio, como fenómeno social y existencial. Subrayan que, al vincular la confianza con la prevención de discriminaciones algorítmicas, se evita la cosificación y alienación de las personas, reforzando la necesidad de una IA educativa alineada con valores de justicia y respeto a la dignidad humana.

La confianza se relaciona con el principio de transparencia, facilita la legitimidad institucional y da a los estudiantes una expectativa razonable y positiva respecto a su derecho a la privacidad.

2. 7. La Confianza y los Principios Éticos de Uso de la IA

Con el propósito de hilvanar todo lo mencionado hasta ahora se propone articular la confianza con algunos de los principios éticos revisados anteriormente. La confianza puede considerarse el valor que hace operativos y sostenibles los principios éticos de uso de IA en la educación.

Es muy clara la relación entre la explicabilidad y la confianza. Como señala Rosales Gallego (2024), la transparencia y la posibilidad de entender cómo funcionan los algoritmos refuerzan la confianza de los usuarios.

En cuanto a la transparencia y responsabilidad, Pasetti y de Oliveira (2024) integran la confianza dentro de la seguridad existencial de la IA y afirman que la confianza en los

sistemas algoritmos educativos sólo pueden sostenerse si se garantiza la transparencia en cómo funcionan los algoritmos (por ejemplo, explicando sus criterios de decisión) y si se asegura la responsabilidad de los actores humanos que supervisan esas decisiones. En este sentido es importante la propuesta de ACE que menciona el co-diseño de los docentes y por lo tanto la co-responsabilidad.

La confianza se ve potenciada cuando los sistemas garantizan la supervisión humana, que es otro principio ético, de modo que los docentes puedan intervenir, corregir o auditar procesos algorítmicos sin quedar relegados (Rosales Gallego, 2024). Aplicado al ámbito educativo, esto implica que la tecnología no sustituye el juicio pedagógico, sino que lo amplifica (ACE).

Si los sistemas de IA ofrecen igualdad de trato y no producen sesgos, aumenta la confianza de los usuarios en que sus intereses serán respetados (Rosales Gallego, 2024), en otras palabras, la confianza se vincula con la percepción de equidad.

Comprender el funcionamiento y los límites de la IA fomenta la confianza informada. Desde la perspectiva conectivista (Castañeda 2013), la alfabetización crítica permite a docentes y estudiantes involucrarse de forma más activa y confiada en el uso de tecnologías educativas. La confianza se conecta directamente con la privacidad, en atención a que los usuarios solo compartirán datos sensibles si perciben que existen garantías de protección, confidencialidad y control. El Grupo de Expertos de Alton Nivel en IA (2019) sitúa la solidez técnica y la seguridad entre los siete requisitos de una IA fiable, y el NIST (2023) las incorpora como características de confiabilidad verificables.

En definitiva, el estudiante tendrá confianza si es oportunamente informado y comprende qué datos se recaban, con qué finalidades, durante cuánto tiempo, con quién se comparten y qué consecuencias puede tener su uso, especialmente cuando intervienen sistemas de IA (Jones, 2019). A partir de esta información y comprensión podrá otorgar un consentimiento informado con base en la confianza.

Capítulo 3. El Consentimiento Informado del Estudiante. Comparación de los Sistemas de la UE, EUA y México.

A diario encontramos millones de aceptaciones para el tratamiento de datos que revelan estados de nuestra vida privada con la única finalidad de gozar de aplicaciones que nos hagan más cómoda la vida pero que en realidad suponen un ejercicio profundo de penetración en nuestros pensamientos, sentimientos y reacciones más profundas. Ante esta realidad ¿habrá forma de proteger en algo nuestra privacidad? (Tenorio, 2021, p. 127)

La aplicación de la IA optimiza tanto la eficiencia administrativa como la educativa, permitiendo una mejora en la gestión de contenidos y el seguimiento del progreso de los estudiantes (Holmes, Bialik, & Fadel, 2019). Sin embargo, también plantea importantes desafíos éticos y jurídicos, por lo que es necesario establecer un marco ético y jurídico claro, inclusivo y fundamentado en principios sólidos que consideren a todos los actores involucrados en el ecosistema educativo.

Dentro de estos principios centrales para regular la IA en la educación, encontramos que el respeto a la privacidad del estudiante es un aspecto donde la ética y el derecho convergen. Convergen porque tanto la ética como el derecho consideran fundamental proteger la privacidad de los datos de los estudiantes, consagrando la privacidad como un principio ético y como un derecho regulado.

Desde una perspectiva general, la preocupación y el interés por la privacidad es un tema que ha sido ampliamente abordado en la cultura occidental para comprenderse y abordarse desde diversos campos de estudio. Estas incluyen disciplinas clásicas como el derecho y la filosofía (Ed. Schoeman, 1984; Fried, 1968), y otras más recientes que se centran específicamente en la privacidad de la información como la ética y la informática (Floridi, 2006; Mittelstadt et al., 2016; Nissenbaum, 2011). A medida que avanza este milenio, con la irrupción de la IA en tantos aspectos cotidianos, se cuestiona cómo afecta esta tecnología a la privacidad de la persona (Mittelstadt et al., 2016; Garfinkel & Bowen, 2022).

3. 1. Derecho a la Privacidad en la era de la IA

La privacidad es un concepto dinámico y en evolución constante (Quijano Decanini, 2022). La doctrina ha clasificado las distintas acepciones de este derecho en seis grandes rubros: “1. El derecho a ser dejado solo, 2. El ámbito de acceso limitado a uno mismo, 3. La facultad de ocultar ciertos asuntos, 4. El control de la información personal, 5. La protección de la personalidad, individualidad y la dignidad y, 6. El derecho a la intimidad o el control sobre las relaciones y aspectos de la vida” (Solove, 2010 Citado por Quijano Decanini p. 55).

Asimismo, la privacidad se reconoce como un derecho humano y un presupuesto para el libre desarrollo de la personalidad, la dignidad y la autonomía, tanto en entornos analógicos como digitales (Naciones Unidas, 1948;1966; Organización de los Estados Americanos [OEA], 1969).

Diversos tratados internacionales han subrayado la importancia de proteger este derecho. Por ejemplo, el artículo 12 de la Declaración Universal de Derechos Humanos, el artículo 17 del Pacto Internacional de Derechos Civiles y Políticos, el artículo 16 de la Convención sobre los Derechos del Niño, el artículo 14 de la Convención Internacional sobre la Protección de los Derechos de Todos los Trabajadores Migratorios y de sus Familiares, el artículo 22 de la Convención sobre los Derechos de las Personas con Discapacidad, el artículo 10 de la Carta Africana sobre los Derechos y el Bienestar del Niño, el artículo 11 de la Convención Americana sobre Derechos Humanos (ONU, 2018) y el artículo 8 del Convenio para la Protección de los Derechos Humanos y de las Libertades Fundamentales (Convenio Europeo de Derechos Humanos) destacan la privacidad como un derecho humano esencial, ligado a la dignidad humana y la protección de la autonomía y la identidad personal (Alta Comisionada de Naciones Unidas, 2021).

A nivel global, se observa que se han desarrollado dos enfoques predominantes en lo que respecta al tema de protección de los datos personales. Por un lado, el modelo europeo que prioriza salvaguardar la información y la propiedad de esta, con la finalidad de preservar la dignidad y honorabilidad del individuo, incluso después de su fallecimiento. Este enfoque se fundamenta en la protección de los derechos humanos. Por otro lado, se encuentra el modelo estadounidense, que concibe la protección de los datos personales desde la perspectiva del derecho a la privacidad, el cual puede extinguirse tras la muerte del titular. Este modelo ha surgido principalmente por razones comerciales, dado que las empresas han hecho un uso indiscriminado de la información personal (Sánchez y Rojas, 2012, p. 4).

Con el auge de la digitalización y el manejo constante de información personal, proteger el derecho a la privacidad se ha convertido en una necesidad crítica. El entorno digital en el que vivimos, junto con el desarrollo acelerado de la ciencia de datos (*Big Data*) y la IA, ha transformado muchos aspectos de nuestra vida, incluyendo la privacidad sobre los datos personales. El contexto, el alcance y la naturaleza de los datos disponibles han cambiado drásticamente (Slade & Prinsloo, 2013). A continuación, se explican estos conceptos.

El contexto hace referencia a la Cuarta Revolución Industrial, que, como señalamos en el capítulo anterior, tiene la particularidad de conectar el mundo físico con el mundo digital. A través del internet de las cosas, la información puede ser recopilada, analizada y utilizada no solo por personas, sino también por algoritmos. Esto ocurre mediante las nuevas técnicas de análisis de BD y de IA.

En cuanto a la naturaleza de los datos, nos referimos a información sustantiva y a los metadatos. Estos datos proporcionan información sobre el comportamiento, las relaciones sociales, las preferencias y la identidad de una persona (Porcelli & Martínez, 2020). A lo anterior se suma que el alcance de las tecnologías del BD e IA es cada vez más poderoso, y permiten no sólo obtener y analizar datos cada vez más precisos de las personas, sino predecir, la toma de decisiones, frecuentemente automatizadas, e incluso manipular los comportamientos de los individuos.

En cuanto al cambio en la generación de los datos, se considera que hay una generación consciente e inconsciente. Así lo señala la Asociación por los Derechos Civiles en su informe “El sistema de protección de datos personales en América Latina: Oportunidades y desafíos para los derechos humanos”, en el que destaca que los datos de un individuo pueden ser generados de manera consciente (aquellos que desea compartir) y también inconscientemente, a través de cada movimiento que realiza en línea, generalmente sin que el usuario sea consciente de ello o tenga control sobre dicha generación (Asociación por los Derechos Civiles, 2016).

¿Por qué tanto interés en los datos personales inconscientes y conscientes? Para responder a esta cuestión, se puede aplicar el trillado principio de pesquisas: sigue el rastro (*follow the money*).

Los datos personales cuentan con un valor económico, equiparable a ciertos activos intangibles, tales como el software o el valor comercial de los nombres de dominio (Mendoza Enríquez, 2018). Esto ha llevado a considerarlos como el petróleo de la sociedad de la información y del conocimiento: *data is the new oil* (la información es el nuevo petróleo) señaló *The Economist* en un artículo del 2017.

En la era del BD los datos tienen un carácter dual: por una parte, son un activo esencial en la economía digital y, por otra, son un bien jurídico tutelado por el derecho a la privacidad (Lynskey & Costa-Cabral, 2017). En cuanto a los primeros aspectos varios autores señalan la creciente comercialización de la información personal (Moreno Pérez y Olmeda García, 2021) y su establecimiento como un recurso de alto valor (Sánchez Díaz, 2024); en relación con lo segundo, existe en los usuarios del mundo digital una falta de conciencia sobre sus propios derechos (Moreno Pérez y Olmeda García, 2021).

Por lo anterior, como señala Porcelli “resulta necesario actualizar el significado del derecho a la privacidad en una economía digital” (2019 p. 465), esta resignificación del derecho a la privacidad ha llevado a proponer nuevos derechos digitales que se derivan de esta, tales como: el derecho al olvido digital, como un paradigma emergente que permite a los usuarios gestionar su identidad digital y la permanencia de sus datos en la red (Álvarez Caro, 2015), el derecho de oponerse a la elaboración de perfiles, el derecho a la portabilidad de datos, la autodeterminación informativa y el derecho a la neutralidad del Internet. Estos derechos, para

hacerse efectivos, implican una coordinación internacional de las normas y los agentes implicados, porque el ciberespacio trasciende las fronteras (Quijano Decanini 2022; Tenorio Cueto, 2021).

En esencia, la privacidad implica proteger a la persona de la intrusión, interferencia y acceso a su información por parte de otros (Alta Comisionada de Naciones Unidas, 2021). Es decir, una especie de control individual sobre la información que desea compartir. Este enfoque de control de la privacidad no significa que la información y los datos personales no sean accesibles a otros, más bien hace referencia a que la persona puede decidir quién puede acceder a su información y en qué condiciones (Fried, 1968; Froomkin, 2000; Nissenbaum, 2011).

Como afirma Jones “la privacidad como control se inclina hacia la elección individual” (2019, p. 5), y se manifiesta en el consentimiento. Los marcos jurídicos de todo el mundo siguen haciendo hincapié en el consentimiento, o control individual, como principio fundamental de la legislación sobre privacidad. También, en la doctrina jurídica, la privacidad se asocia con el acceso a la información personal (Nissenbaum, 2011).

Para proteger la privacidad, un elemento clave es el consentimiento informado, tanto desde la perspectiva ética como jurídica (Sedenberg y Hoffmann, 2016). El consentimiento informado es clave, porque apela tanto al entendimiento —a que el sujeto comprenda— como a la voluntad —que el sujeto quiera—. Esta figura jurídica será el tema central de este capítulo. Se expondrá primero en general, y luego centrado en la educación, presentando diversas normativas internacionales y realizando un análisis comparativo con la legislación nacional.

3. 2. Consentimiento Informado: Origen y Evolución

El concepto de consentimiento y, posteriormente, consentimiento informado comenzó a ser empleado y desarrollado en el ámbito médico y de la bioética (Whitney et al., 2004; Welch et al., 2016; Tunzi et al., 2021; Faiad, 2022).

A partir de los juicios de Núremberg (1947), que se llevaron a cabo tras la Segunda Guerra Mundial, se establecieron nuevas reglas que indicaban que el consentimiento voluntario sólo podía obtenerse libremente, sin coerción, y alertando al paciente sobre los riesgos y peligros del procedimiento específico que se le aplicaría (Prinsloo & Slade, 2018, p. 120). No obstante, en décadas posteriores, el Estudio de Tuskegee (1972), experimento clínico donde no se obtuvo el consentimiento de los participantes, evidenció la necesidad urgente de regular el consentimiento informado en la práctica médica.

El Informe Belmont, publicado en 1979 en Estados Unidos, en respuesta a los abusos cometidos en el Estudio de Tuskegee y otros experimentos, marcó un cambio significativo al definir principios esenciales para la investigación con humanos. Este documento desempeñó

un papel crucial en el desarrollo del concepto de consentimiento informado, tanto en el ámbito médico como en la investigación científica. Los tres principios fundamentales que introdujo —respeto a las personas, beneficencia y justicia— tuvieron un impacto profundo en las normativas posteriores sobre el consentimiento informado. Uno de los aspectos más destacados de este informe es su enfoque en que el consentimiento informado no debe considerarse simplemente un procedimiento administrativo, como la firma de un documento, sino un proceso continuo de diálogo y entendimiento mutuo entre el médico y el paciente (*The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*, 1979).

El consentimiento informado, a diferencia del simple consentimiento, no sólo implica la información necesaria al paciente para obtener el consentimiento, sino también asegurarse de que dicho paciente comprenda dicha información. Este proceso de comprensión requiere un diálogo constante: “un intercambio de ideas entre el paciente y el médico y una colaboración entre ambos para la toma de decisiones” (Whitney et al., 2004, p. 55). En este sentido, la decisión del paciente de otorgar su consentimiento se basa en el consejo del médico.

El consentimiento informado tiene tres elementos: (1) que el paciente comprenda la situación, (2) que la decisión se tome de manera conjunta, y (3) que se base en la confianza del paciente en el médico. El primer elemento —la comprensión— debe ser abordado desde dos perspectivas: la del paciente y la del médico. El paciente debe tener el conocimiento suficiente de la situación que le es explicada, mientras el médico debe ser capaz de evaluar hasta qué punto el paciente ha comprendido dicha información. (Prinsloo & Slade, 2018).

El segundo elemento crucial es la decisión compartida, que se da cuando existen diversas opciones y el médico involucra al paciente en la elección de la opción más adecuada (Whitney et al., 2004, p. 55). Este elemento es fundamental porque demuestra que el consentimiento informado no se reduce a la firma de un formulario, más bien debe ser un proceso de diálogo continuo. Como señala Whitney, el consentimiento informado. “No ocurre cuando se firma un formulario; ocurre cuando el paciente y el médico discuten un problema y eligen juntos una intervención, un proceso que puede tener lugar en una sesión o en el transcurso de varios encuentros” (Whitney et al., 2004, p. 54). Barajas-Ochoa también subraya esta distinción, afirmando que “el consentimiento informado es un proceso de comunicación y el documento de consentimiento informado es solo una forma de plasmar la autorización del paciente. Que un paciente firme el documento no implica la existencia del consentimiento” (Barajas-Ochoa & Barajas-Ochoa, 2019).

El tercer aspecto es la confianza del paciente en el médico para la validez del consentimiento. A medida que el procedimiento se vuelve más complejo y la terminología médica más técnica, la confianza en el profesional de salud es lo que impulsa al paciente en la toma de una decisión, incluso si no comprende completamente todos los datos proporcionados. Esta

confianza se basa en el deber fiduciario de cuidado y en el juramento hipocrático que los médicos realizan al comenzar su práctica profesional.

La Organización de las Naciones Unidas (ONU) ha abordado el tema del consentimiento informado, especialmente en relación con los derechos humanos y la ética en la investigación científica y médica. Este organismo enfatiza la necesidad de obtener el consentimiento de las personas en diversas circunstancias, en particular en la investigación biomédica, los tratamientos médicos y la protección de la información personal (Declaración Universal sobre Bioética y Derechos Humanos, 2005).

A pesar de los esfuerzos regulatorios, se ha observado que el consentimiento informado ha pasado de ser un proceso de toma de decisión conjunta -paciente y médico- a un enfoque centrado en la firma de formularios, lo que ha llevado a la simplificación de la noción original de consentimiento informado.

El tema del consentimiento informado es ampliamente aceptado, incluso fuera del ámbito médico, por ejemplo, en la educación (Jones, 2019) (Tene & Polonetsky, 2013). Ante esta situación surge la pregunta: ¿realmente estamos considerando un consentimiento informado al hablar del formulario de consentimiento? Nos parece que no y expondremos con claridad el porqué.

3. 3. El Consentimiento Informado en la ONU

En la era digital, el consentimiento informado cada vez se considera con mayor importancia dentro de la Organización de las Naciones Unidas (ONU), debido a los avances tecnológicos, la extendida recopilación masiva de datos y la interconexión global. En 2013 la Asamblea General de la ONU adoptó la Resolución A/RES/68/167 “El Derecho a la Privacidad en la Era Digital”; en la cual se afirma que: “la vigilancia y la interceptación ilícitas o arbitrarias de las comunicaciones, así como la recopilación ilícita o arbitraria de datos personales, al constituir actos de intrusión grave, violan los derechos a la privacidad y a la libertad de expresión y pueden ser contrarios a los preceptos de una sociedad democrática”, de igual manera instó a implementar acciones para poner fin a las violaciones del derecho a la privacidad, incluyendo aquellas relacionadas con las comunicaciones digitales, y a realizar revisiones de los sistemas de vigilancia en este ámbito.

Dos años más tarde el Consejo de Derechos Humanos de las Naciones Unidas el 24 de marzo de 2015 adoptó la resolución A/HRC/28/L. 27 sobre “El derecho a la privacidad en la era digital” y creó la figura de un Relator especial sobre el derecho a la privacidad en la era digital, nombrándose al Profesor Joseph A. Cannataci, por un periodo inicial de tres años (ONU, Consejo de Derechos Humanos, 2015). De manera enunciativa, se le solicitó la recopilación de información relevante, la identificación de obstáculos probables, la

participación en iniciativas globales, la sensibilización y la presentación de informes anuales al Consejo de Derechos Humanos y la Asamblea General de las Naciones Unidas.

Asimismo, en el año 2015, Naciones Unidas aprobó la Agenda 2030, un plan de acción internacional compuesto por 17 objetivos, los denominados Objetivos de Desarrollo Sostenible (ODS); y ha sido suscrita por 193 países. Dicha Agenda promovió el desarrollo económico, social y ambiental, con énfasis en la inclusión y el respeto a los derechos humanos.

Aunque el consentimiento informado en el ámbito digital no se abordó de manera directa, la Meta 16.10 hace referencia a garantizar el acceso a la información, a promover las libertades fundamentales, el acceso de la sociedad a la información y la protección de los derechos humanos, de acuerdo con la legislación nacional e internacional (Naciones Unidas, 2015). Este principio está estrechamente vinculado con el derecho a la privacidad y la imperiosa necesidad de proteger los datos personales en un contexto digital.

Continuando en esta línea, el 19 de diciembre de 2016 la Asamblea General de la ONU adoptó la Resolución N° 71/199, titulada “El derecho a la privacidad en la era digital”. Esta resolución reafirma el derecho a la privacidad y expone la situación dicotómica que el entorno actual representaba: por un lado “el rápido ritmo del desarrollo tecnológico permitía a las personas de todo el mundo utilizar las nuevas tecnologías de la información y las comunicaciones y, [por otro] incrementar la capacidad de los gobiernos, las empresas y personas de llevar a cabo actividades de vigilancia, interceptación y recopilación de datos” (Porcelli, 2019). En esta situación, la ONU destacó que muchas veces las personas no daban su consentimiento libre, explícito y fundado a la venta o a la reventa múltiple de sus datos personales (ONU, Asamblea General, 2016).

El Alto Comisionado de las Naciones Unidas (ACNUDH) el 3 de agosto de 2018, emitió el informe “*The right to privacy in the digital age*” (El derecho a la privacidad en la era digital) donde vuelve a señalar la preocupación, basada en evidencias, de que muchos de los datos que se obtienen de una persona se recaban sin su consentimiento válido. El documento enfatiza la responsabilidad del Estado en la creación de normativas y en la supervisión eficaz y efectiva del derecho a la privacidad destacando la relevancia del consentimiento informado. En este contexto, el punto 29 de este documento señala que “para prevenir el uso arbitrario de la información personal, el tratamiento de datos personales debería basarse en el consentimiento libre, específico, informado y sin ambigüedades de las personas interesadas, o en otro fundamento legítimo previsto en la ley” (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2018, p. 10).

Durante los años 2020 y 2021 se han publicado informes por parte del ACNUDH, que si bien han tratado del tema de privacidad no se han centrado en aspectos relativos al consentimiento informado (ACNUDH, 2021).

3. 4. El Consentimiento Informado en la Legislación Europea

La Unión Europea ha hecho el esfuerzo más ambicioso por regular la protección de datos de manera integral, en un proceso de armonización que ha evolucionado a lo largo de décadas (Alsenoy, 2016).

En este apartado se analizarán dos regulaciones fundamentales para esta investigación: el Reglamento General de Protección de Datos (RGPD) de 2016 y el Reglamento de IA de la Unión Europea (AI Act), publicado en 2024. En conjunto, estas normas inciden en la protección de la privacidad y en la transparencia del tratamiento de datos en contextos educativos, mediante un enfoque basado en el riesgo, obligaciones de información y, en supuestos específicos, exigencias de consentimiento u otras garantías reforzadas (Parlamento Europeo y Consejo de la Unión Europea, 2016; Parlamento Europeo y Consejo de la Unión Europea, 2024).

En esta investigación no se analizan las normas particulares de los Estados miembros en materia educativa, ya que el foco se sitúa en los estándares mínimos comunes derivados del RGPD y del AI Act. No obstante, se mencionarán algunos ejemplos de aplicación centrados en el caso español, por tratarse del ejemplo elegido para esta investigación.

3. 4. 1. El Reglamento General de Protección de Datos (RGPD)

La entrada en vigor del RGPD constituyó un hito clave en la salvaguarda de los derechos fundamentales en el entorno digital, además de ofrecer un marco regulatorio preciso para las entidades empresariales y los organismos públicos en el mercado único digital. El Parlamento Europeo, como órgano legislativo de la UE desempeñó un papel fundamental para la armonización de la legislación en el ámbito europeo, permitió reducir la fragmentación entre los distintos regímenes nacionales y disminuyó las cargas administrativas innecesarias, favoreciendo así una mayor certeza jurídica y optimización en la gestión de datos personales.

El RGPD fue aprobado en 2016, entró en vigor 20 días después y fue aplicado a partir del 25 de mayo de 2018. El RGPD prioriza la gestión de las violaciones de datos, entendidas como cualquier incidente de seguridad que afecte la confidencialidad, integridad o disponibilidad de la información personal.

Si bien el consentimiento es una de las bases jurídicas más conocidas para el tratamiento de datos personales, no constituye la única fundamentación legal. El RGPD establece otras cinco bases legales. Como señala Lama Muñoz “La base legitimadora del tratamiento comprenderá las distintas etapas del ciclo de vida de una IA, pudiendo ser la misma o distinta. Con carácter general, según el artículo 6 del RGPD se podrá justificar el tratamiento en ejecución de un contrato, interés legítimo o el consentimiento de los interesados” (Lama Muñoz, 2025, p. 6).

No obstante, el RGPD establece una definición jurídica robusta del consentimiento, sentando un estándar elevado para su validez como base legitimadora del tratamiento (Politou, Alepis & Patsakis 2018). El artículo 4. 11 lo define conceptualmente como toda "manifestación de voluntad libre, específica, informada e inequívoca por la cual el interesado acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen" (RGPD, art. 4. 11).

El RGPD detalla que el consentimiento debe expresarse en una declaración afirmativa o una acción clara que evidencie la voluntad libre, específica, informada e inequívoca del interesado, no puede presumirse ni darse de manera implícita. Tal como se establece en el considerando 32, se permite el uso de casillas para obtener el consentimiento, pero se deja claro que una casilla pre-marcada o la falta de acción por parte del interesado no son suficientes para considerar que se ha prestado el consentimiento. Además, se estipula que, si existen varios fines para el tratamiento de los datos, el consentimiento deberá ser obtenido de manera separada para cada uno de esos fines.

Asimismo, para que el consentimiento sea verdaderamente informado deben concurrir, al menos, la capacidad para consentir, la voluntariedad, la provisión de información suficiente y la comprensión por parte del titular (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). La consideración 42 del RGPD, establece que “para que el consentimiento sea informado, el interesado debe conocer como mínimo la identidad del responsable del tratamiento y los fines del tratamiento a los cuales están destinados los datos personales”.

La transparencia en la recopilación y el uso de los datos personales es un requisito esencial para garantizar el ejercicio del derecho a la protección de datos (Quijano Decanini 2022). Por ello, el RGPD garantiza el derecho de las personas a recibir información clara y transparente sobre la recopilación y el uso de sus datos, estableciendo así diversas obligaciones para el responsable del tratamiento (Tamayo Velasco, 2021). Además de estos requisitos mínimos es conveniente facilitarle al interesado información sobre cómo se van a tratar esos datos, el plazo de conservación, los posibles destinatarios de los datos, las posibles transferencias internacionales de datos y los derechos de los que es titular el interesado.

La finalidad del uso o tratamiento de datos es crucial en el ámbito del consentimiento. El responsable debe especificarla claramente. Si el responsable desea utilizar los datos personales del interesado para un propósito diferente al original, deberá obtener un nuevo consentimiento para cada finalidad específica. Como señala González Guerrero “La granularidad es otra condición fundamental. Si se recolecta información para distintos propósitos, la persona debe poder aceptarlos y rechazarlos selectivamente. Si no se busca un consentimiento separado para cada propósito, existe una falta de libertad” (2018, p. 226).

El RGPD distingue entre dos situaciones: cuando los datos personales son obtenidos directamente del titular (artículo 13 del RGPD) y cuando la información se recopila a través de terceros (artículo 14 del RGPD). Es importante señalar que el responsable del tratamiento de los datos puede ser distinto del que recopiló los datos.

Aunque ambas disposiciones son muy similares en cuanto a la obligación impuesta al responsable del tratamiento de datos, difieren en relación con el origen de la información. El artículo 14 del RGPD aborda el caso en que la información personal no se obtiene del interesado directamente sino a través de otro que es quien recaba la información. En consecuencia, impone una carga adicional al responsable del tratamiento quien, además de recibir la información de un tercero, debe establecer una relación con el interesado, para informarle del tratamiento de sus datos (Comandé, 2022).

Además de establecer el concepto de consentimiento libre, el RGPD en su artículo 7 detalla las condiciones que deben cumplirse para su validez: establece las condiciones para la demostración (7. 1) y obtención del consentimiento (7. 2) y regula la posibilidad de que el individuo retire dicho consentimiento en cualquier momento (7. 3).

Al evaluar si el consentimiento se ha dado libremente, se tendrá en cuenta en la mayor medida posible el hecho de si, entre otras cosas, la ejecución de un contrato, incluida la prestación de un servicio, se supedita el consentimiento al tratamiento de datos personales que no son necesarios para la ejecución de dicho contrato.

3. 4. 2. Directrices 5/2020 sobre el Consentimiento Informado

Para detallar aún más el tema del consentimiento informado se emitieron las Directrices 5/2020 sobre el consentimiento del Comité Europeo de Protección de Datos (CEPD). Este documento es una guía interpretativa para definir cómo se debe interpretar y aplicar el consentimiento como base jurídica en el Reglamento (UE) 2016/679 (RGPD).

El corazón de las directrices gira en torno a las condiciones del consentimiento válido tal y como lo define el RGPD: una manifestación de voluntad libre, específica, informada e inequívoca del interesado ya sea mediante una declaración o una acción afirmativa clara.

El CEPD reitera que el consentimiento válido requiere elección y control reales; si el individuo no puede rechazar o retirar sin consecuencias negativas, el consentimiento no es libre (Comité Europeo de Protección de Datos, 2020). Además, se refuerza la obligación de granularidad por finalidades, de modo que, si para una misma recogida hay múltiples finalidades, se debe permitir elegir independientemente por cada una de ellas, y si surge una nueva finalidad, se debe recabar un nuevo consentimiento, a menos que exista otra base jurídica más apropiada. (CEPD, 2020).

Las directrices concretan estos principios en obligaciones auditables: el consentimiento no puede deducirse del silencio o la inacción, ni de casillas premarcadas, ni es suficiente seguir utilizando un servicio como señal de aceptación (CEPD, 2020).

En las directrices se subraya la obligación del responsable de demostrar que obtuvo el consentimiento, manteniendo registros sobre cómo y cuándo se recabó y qué información se facilitó al interesado en ese momento. Es necesario que la retirada del consentimiento sea posible en cualquier momento y que sea tan sencillo como darlo, preferiblemente por la misma vía en que se dio cuando se utilizaron medios electrónicos, sin perjuicio y de forma gratuita, y una vez retirada la autorización, el responsable tiene que suspender el tratamiento que se fundamentaba en ella (CEPD, 2020).

3. 4. 3. Ley de la IA (AI Act)

Asimismo, el Reglamento (UE) 2024/1689, llamada Ley de inteligencia artificial (AI Act) entró en vigor el 1 de agosto de 2024, pero su aplicación es progresiva. El plan original establece el siguiente cronograma: desde el 2 de febrero de 2025 resultan aplicables, entre otras, las disposiciones sobre prácticas prohibidas y la alfabetización en IA; desde el 2 de agosto de 2025 se activan, entre otros extremos, reglas de gobernanza y obligaciones para modelos de IA de propósito general; desde el 2 de agosto de 2026 aplica la mayor parte del Reglamento, incluyendo obligaciones para sistemas de alto riesgo (Anexo III) y deberes de transparencia; y, finalmente, desde el 2 de agosto de 2027 aplicarán obligaciones específicas para IA de alto riesgo incorporada en productos regulados. En consecuencia, el análisis jurídico de esta investigación distingue entre obligaciones ya exigibles y obligaciones de exigibilidad diferida, evitando inferencias de “cumplimiento pleno” antes de las fechas de aplicación correspondientes.

El objetivo principal de esta ley es promover la adopción de sistemas de IA que estén orientados hacia el bienestar humano y sean confiables, asegurando a su vez un alto nivel de protección en cuanto a la salud, la seguridad y los derechos fundamentales establecidos en la Carta de los Derechos Fundamentales de la Unión Europea.

Esta ley tiene especial relevancia para el sector educativo, porque en la misma se le considera especialmente, y para mostrarlo, se transcribe íntegro el considerando 56 de la AI Act:

El despliegue de sistemas de IA en la educación es importante para promover una educación y formación digitales de alta calidad y para permitir que todos los alumnos y profesores adquieran y compartan las habilidades y competencias digitales necesarias, incluida la alfabetización mediática, y el pensamiento crítico, para participar activamente en la economía, la sociedad y los procesos democráticos. Sin embargo, los sistemas de IA utilizados en la educación o la formación profesional, en particular para determinar el

acceso o la admisión, para asignar personas a instituciones o programas educativos y de formación profesional a todos los niveles, para evaluar los resultados del aprendizaje de las personas, para evaluar el nivel de educación adecuado para una persona e influir materialmente en el nivel de educación y formación que las personas recibirán o al que podrán acceder, o para supervisar y detectar comportamientos prohibidos de los estudiantes durante las pruebas, deben clasificarse como sistemas de IA de alto riesgo, ya que pueden determinar el curso educativo y profesional de la vida de una persona y, por lo tanto, pueden afectar a la capacidad de esa persona para asegurarse un medio de vida. Cuando se diseñan y utilizan de forma inadecuada, estos sistemas pueden ser especialmente intrusivos y vulnerar el derecho a la educación y a la formación, así como el derecho a no ser discriminado, y perpetuar pautas históricas de discriminación, por ejemplo, contra las mujeres, determinados grupos de edad, personas con discapacidad o personas de determinados orígenes raciales o étnicos u orientación sexual.

En el artículo 3 de esta normativa, se da la definición de consentimiento informado como la manifestación libre, específica, inequívoca y voluntaria de un sujeto de su voluntad de participar en un ensayo concreto en condiciones reales, tras haber sido informado de todos los aspectos del ensayo que son relevantes para la decisión del sujeto de participar.

Para comprender el contexto del consentimiento informado aplicado a la educación es importante saber que la AI Act se centra en un enfoque basado en riesgo y propone que los sistemas de IA que puedan utilizarse en distintas aplicaciones se analicen y clasifiquen según el riesgo que supongan para los usuarios. Los distintos niveles de peligro implicarán una mayor o menor regulación (Parlamento Europeo y Consejo de la Unión Europea, 2024).

El AI Act adopta un enfoque basado en el riesgo y establece obligaciones diferenciadas según el tipo de sistema, incluyendo prácticas prohibidas, sistemas de alto riesgo y deberes específicos de transparencia para ciertos usos (Parlamento Europeo y Consejo de la Unión Europea, 2024). Éstas se clasifican en los siguientes riesgos:

Riesgo inaceptable: Sistemas prohibidos que amenazan la seguridad, derechos o valores democráticos, como la puntuación social gubernamental, manipulación cognitivo-conductual dañina y vigilancia biométrica masiva indiscriminada en espacios públicos.

Alto riesgo: IA en áreas críticas como sanidad, educación, empleo, aplicación de la ley, justicia, infraestructuras críticas y servicios esenciales. Requiere pruebas estrictas,

transparencia, supervisión humana, sistema de gestión de riesgos, calidad de datos, documentación técnica, ciberseguridad, monitorización y marcado CE².

Riesgo limitado: Incluye chatbots y asistentes virtuales. Los usuarios deben estar informados cuando interactúen con IA, pero con menos normas.

Riesgo mínimo: Aplicaciones de bajo riesgo, como filtros de spam o IA de juegos, en gran medida exentas de regulación.

A continuación, se presenta un gráfico para ilustrar estos tipos de riesgo. El sector educativo se encuentra contemplado en el riesgo alto según el anexo III número 3 y el considerando 56.

Figura 8. Tipos de riesgo según la AI Act de la UE.



Nota. Adaptada de An European strategy for artificial intelligence [Presentación], por L. Sioli, 2021, Comisión Europea. Traducción propia. (https://cdn.gelestatic.it/repubblica/blogautore/sites/408/2023/02/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21_compressed.pdf)

La referencia específica al consentimiento informado se da en el marco del riesgo de alto nivel, dentro del capítulo VI de la IA Act titulado “Medidas de apoyo a la innovación”. En el artículo 61 se establece que, para las pruebas de sistemas de IA de alto riesgo en condiciones del mundo real fuera de espacios aislados de regulación de la IA, es necesario que antes de las pruebas se obtenga el consentimiento informado de los sujetos participantes, detallando que éste sea dado libremente antes de participar en dichas pruebas y después de haber recibido información concisa, clara, pertinente y comprensible.

² Es el sello oficial de conformidad de la UE que demuestra que un sistema de IA de alto riesgo cumple con todas las normas de seguridad, protección de datos y gestión de riesgos exigidas por la ley comunitaria.

Para garantizar que sea debidamente informado con datos concisos, claros, pertinentes y comprensibles, se detalla el contenido de éstos:

- (a) la naturaleza y los objetivos de las pruebas en condiciones reales y los posibles inconvenientes que puedan derivarse de su participación;
- (b) las condiciones en las que se realizarán los ensayos en condiciones reales, incluida la duración prevista de la participación del sujeto o sujetos;
- (c) sus derechos y las garantías relativas a su participación, en particular su derecho a negarse a participar en ensayos en condiciones reales y a retirarse de los mismos en cualquier momento, sin que ello suponga perjuicio alguno y sin tener que aportar justificación alguna;
- (d) las modalidades para solicitar la anulación o el incumplimiento de las predicciones, recomendaciones o decisiones del sistema de IA;
- (e) el número único de identificación a escala de la unión del ensayo en condiciones reales, de conformidad con el artículo 60 apartado 4, letra c), y los datos de contacto del proveedor o de su representante legal de quien pueda obtenerse más información.

Así se establece que el consentimiento informado se fechará y documentará y se entregará una copia a los sujetos de ensayo o sus representantes legales. (AI Act Art. 61.)

Del artículo anterior podemos deducir que el consentimiento informado se identifica con el documento o el aviso de consentimiento y se acota al ensayo o pruebas consideradas de alto riesgo.

Si continuamos analizando la AI Act, en el anexo III encontramos que se detalla de modo concreto algunos sistemas de IA que se utilizan en la educación y formación profesional. Esta enumeración presenta un aspecto positivo, pero corre el riesgo de no ser actualizado en un futuro, porque los sistemas de IA están en constante evolución.

Anexo III de AI Act. Sistemas de IA que se utilizan en la educación y formación.

- (a) Sistemas de IA destinados a ser utilizados para determinar el acceso o la admisión o para asignar personas físicas a instituciones educativas y de formación profesional de todos los niveles;
- (b) Sistemas de IA destinados a ser utilizados para evaluar los resultados del aprendizaje, incluso cuando dichos resultados se utilizan para dirigir el proceso de aprendizaje de las personas físicas en centros educativos y de formación profesional de todos los niveles;

- (c) Sistemas de IA destinados a ser utilizados con el fin de evaluar el nivel adecuado de educación que un individuo recibirá o al que podrá acceder, en el contexto de o dentro de instituciones educativas y de formación profesional a todos los niveles;
- (d) Sistemas de IA destinados a ser utilizados para controlar y detectar comportamientos prohibidos de los estudiantes durante las pruebas en el contexto de las instituciones educativas y de formación profesional a todos los niveles o dentro de ellas.

Como podemos observar los sistemas de IA considerados se refieren a temas de admisión, de evaluación de conocimientos y a temas de integridad académica.

En consecuencia, el AI Act prevé la figura del consentimiento informado en supuestos delimitados —por ejemplo, en ensayos en condiciones reales (*real-world testing*) de sistemas de alto riesgo— y, por otra parte, impone deberes de información y transparencia cuando las personas interactúan con determinados sistemas de IA. No obstante, la interacción ordinaria de estudiantes con herramientas de IA en el ámbito universitario se rige principalmente por las bases de licitud, transparencia y garantías del RGPD, así como por las políticas internas de las IES (Parlamento Europeo y Consejo de la Unión Europea, 2016; Parlamento Europeo y Consejo de la Unión Europea, 2024).

La AI Act además de contemplar la figura del consentimiento informado, hace mención expresa en el considerando 56 a la alfabetización mediática y el pensamiento crítico en el uso de IA por parte de los estudiantes y docentes.

La llamada AI Act acaba de ser modificada recientemente, creándose el Reglamento (UE) 2026/1744, en vigor desde el 27 de julio de 2026. En el siguiente inciso se expone este tema.

3. 4. 4. El Paquete Digital Omnibus 2026

El 19 de noviembre de 2025, la Comisión Europea presentó el paquete Digital Omnibus, una iniciativa de simplificación normativa que proponía reabrir simultáneamente el Reglamento (UE) 2016/679, la Directiva 2002/58/CE sobre la privacidad y las comunicaciones electrónicas, la Ley de Datos y el Reglamento (UE) 2024/1689. Conviene precisar, desde el inicio, que el paquete se tramitó como dos expedientes legislativos diferenciados, con destinos y calendarios distintos, circunstancia que resulta determinante para valorar su alcance.

Las reacciones institucionales fueron severas: el Comité Europeo de Protección de Datos y el Supervisor Europeo de Protección de Datos, en su Dictamen conjunto 2/2026, advirtieron que la redefinición del concepto de dato personal y la flexibilización del tratamiento en contextos de IA podían reducir materialmente el nivel de protección vigente (EDPB & EDPS, 2026). La organización no gubernamental de Derechos Humanos Europea NOYB -que impulsa demandas colectivas y casos estratégicos contra grandes empresas tecnológicas que

infringen las normativas de privacidad- sostuvo que diversas enmiendas entrarían en conflicto con la Carta de los Derechos Fundamentales de la Unión Europea y se apartarían de la jurisprudencia del Tribunal de Justicia (NOYB, 2025). La red de Derechos Digitales Europeos (EDRi, 2026) caracterizó el efecto agregado como un desplazamiento estructural. El Reglamento General de Protección de Datos dejaría de operar como marco preventivo centrado en derechos para pasar a depender de autoevaluaciones internas y justificaciones posteriores. Esto implicaría la consiguiente traslación de la carga desde el responsable del tratamiento hacia el titular de los datos.

El expediente relativo a la IA culminó su tramitación en el Reglamento (UE) 2026/1744, de 8 de julio de 2026. Este fue aprobado por el Parlamento Europeo el 16 de junio y por el Consejo, el 29 de junio. Fue publicado en el Diario Oficial de la Unión Europea el 24 de julio y está en vigor desde el 27 de julio de 2026. La norma modifica la Ley de IA junto con los reglamentos sobre aviación civil y sobre máquinas. Su contenido principal consiste en un nuevo calendario de aplicación. Las obligaciones relativas a los sistemas de alto riesgo del anexo III se aplazan al 2 de diciembre de 2027 y las correspondientes al anexo I al 2 de agosto de 2028. Importa subrayar que las obligaciones de transparencia del artículo 50 no fueron aplazadas y resultan aplicables desde el 2 de agosto de 2026, así como que el plazo para el marcado de contenidos generados o manipulados artificialmente se redujo de seis a tres meses. El Reglamento incorpora asimismo dos nuevas prohibiciones al artículo 5, referidas a la generación de material de abuso sexual infantil y de imágenes íntimas no consentidas, y mantiene el umbral de estricta necesidad del artículo 10.5 frente a las propuestas de flexibilización que se habían planteado durante la negociación.

El segundo expediente, que afecta al RGPD, a la normativa sobre privacidad en las comunicaciones electrónicas y a la Ley de Datos, permanecía en tramitación al cierre de esta investigación, con previsión de acuerdo en el primer semestre de 2027. En su discusión, la posición del Consejo retiró algunas de las modificaciones más cuestionadas —singularmente el debilitamiento del artículo 22, que constituye la salvaguarda frente a las decisiones basadas únicamente en tratamiento automatizado, y la ampliación de la definición de investigación científica—, si bien se trata de posiciones provisionales sobre un texto aún no adoptado (EDRi, 2026).

La secuencia resulta significativa para el argumento que aquí se sostiene: el régimen de la IA se flexibilizó antes de que se resolviera el de protección de datos. Ello confirma que la tutela de la privacidad frente a los sistemas de IA en el ordenamiento europeo depende menos de lo que dispone la Ley de IA que de la integridad del Reglamento General de Protección de Datos.

De este marco se derivan tres implicaciones directas para las IES:

1. La calificación como sistemas de alto riesgo. El anexo III, punto 3, de la Ley de IA clasifica como de alto riesgo los sistemas destinados a determinar el acceso o la

admisión de personas a instituciones educativas y de formación profesional, a evaluar los resultados del aprendizaje —incluidos aquellos empleados para orientar el proceso formativo—, a valorar el nivel educativo adecuado para una persona y a supervisar y detectar comportamientos prohibidos durante las evaluaciones. Una institución que despliegue estas herramientas queda sujeta al régimen más exigente del Reglamento, aplicable desde el 2 de diciembre de 2027 conforme al calendario resultante del Reglamento (UE) 2026/1744.

2. La prohibición del reconocimiento de emociones en el ámbito educativo. El artículo 5.1.f) prohíbe la introducción en el mercado, la puesta en servicio y el uso de sistemas de IA destinados a inferir emociones de una persona física en los ámbitos del lugar de trabajo y de los centros educativos, salvo cuando su uso se destine a fines médicos o de seguridad. La excepción es objetiva y de interpretación estricta: no comprende la finalidad pedagógica, ni la evaluación del compromiso o la atención del estudiantado, ni la analítica del aprendizaje. Y, lo que resulta determinante para este apartado, la prohibición no es disponible por las partes: ninguna política institucional de consentimiento, ni la firma de un aviso de privacidad por parte del estudiante, puede habilitar válidamente ese despliegue.
3. La insuficiencia estructural del consentimiento en relaciones asimétricas. El fundamento de la prohibición anterior es explícito y trasciende su supuesto concreto: en relaciones caracterizadas por la asimetría de poder —docente y estudiante, empleador y trabajador— el consentimiento no puede reputarse libre en el sentido del artículo 4.11 del Reglamento General de Protección de Datos. La consecuencia institucional es de calado. Un marco universitario que haga descansar la protección de la privacidad sobre la aceptación individual de un aviso está adoptando precisamente la técnica que el legislador europeo consideró inoperante en este contexto. La alternativa que ofrece la Ley de IA consiste en prohibir determinados usos con independencia de la voluntad de la persona afectada y en imponer requisitos de diseño y de gobernanza a quien despliega el sistema. La garantía, por tanto, debe trasladarse desde la manifestación de voluntad del estudiante hacia la contratación con proveedores, la evaluación previa del sistema y la decisión institucional sobre su despliegue.

3. 4. 5. La Doctrina Especializada

El Reglamento (UE) 2016/679 supuso una evolución sustantiva en la conceptualización del consentimiento. Frente al régimen precedente, exige una manifestación libre, específica, informada e inequívoca mediante declaración o acción afirmativa clara (art. 4.11), lo que excluye el consentimiento tácito, y desplaza hacia el responsable del tratamiento la carga de acreditar que fue efectivamente otorgado (art. 7.1). No se configura, pues, como un requisito formal, sino como condición de validez del tratamiento.

Politou et al. (2018) subrayan esta dimensión material al sostener que "el consentimiento informado se ha otorgado con base en una clara apreciación y comprensión de los hechos, las implicaciones y las consecuencias de una acción" (p. 5). Desde esa premisa, los autores reivindican el consentimiento informado como el último mecanismo de defensa del individuo

frente a la pérdida de autonomía sobre el tratamiento de su información, de modo que cualquier atenuación de la exigencia debería someterse a debate público y a escrutinio crítico si se pretende preservar los ideales democráticos.

La Ley de IA, sin embargo, no prolonga esa lógica. Veale y Zuiderveen Borgesius (2021) advirtieron, en la fase de propuesta, que el texto europeo no responde al modelo de los derechos individuales, sino al de la seguridad de productos: la protección se articula mediante requisitos técnicos verificados *ex ante* por el propio proveedor y acreditados mediante marcado CE, sin intervención alguna de la persona afectada. Holtz y Ledendal (2026) formulan la consecuencia con precisión: dentro del ámbito de aplicación de la Ley de IA, el consentimiento no desempeña ninguna función autónoma, porque la licitud del tratamiento de datos personales continúa rigiéndose íntegramente por el Reglamento General de Protección de Datos. La distribución resultante es nítida: la Ley de IA disciplina cómo se diseña, entrena y despliega el sistema; el Reglamento General de Protección de Datos determina si el dato puede tratarse y bajo qué base jurídica. Las obligaciones de transparencia que sí incorpora la primera —advertir de la interacción con un sistema automatizado y etiquetar los contenidos sintéticos, conforme al artículo 50— constituyen deberes de información y no mecanismos de autorización: informar no equivale a recabar permiso.

Esta separación ha sido objeto de crítica sostenida por las autoridades europeas de protección de datos. En su Dictamen conjunto 5/2021, el Comité Europeo de Protección de Datos y el Supervisor Europeo reclamaron una prohibición general y sin excepciones del reconocimiento biométrico remoto en espacios de acceso público, la proscripción de la categorización biométrica por origen étnico, género u orientación política o sexual, la prohibición de la inferencia de emociones salvo supuestos sanitarios acotados, y la atribución de la función supervisora a las propias autoridades de protección de datos. Advirtieron, además, que la delimitación de las prácticas prohibidas resultaba tan restrictiva que podía quedar reducida a una declaración retórica (EDPB & EDPS, 2021). El texto finalmente aprobado incorporó esas objeciones solo parcialmente.

El punto de mayor tensión doctrinal se localiza en el artículo 10.5, que autoriza a los proveedores de sistemas de alto riesgo a tratar categorías especiales de datos cuando resulte estrictamente necesario para detectar y corregir sesgos. La disposición responde a una paradoja genuina —verificar que un sistema no discrimina exige procesar precisamente los atributos protegidos—, pero habilita el tratamiento de datos sensibles sin intervención del titular. Van Bekkum (2025) ha mostrado que las condiciones acumulativas que la acompañan reproducen principios ya presentes en el Reglamento General de Protección de Datos, y permanece abierta la cuestión de si el precepto crea una base jurídica autónoma o se limita a levantar la prohibición del artículo 9 remitiendo implícitamente a este.

Existe, no obstante, una lectura que interpreta el desplazamiento no como carencia, sino como corrección deliberada. Solove (2013) documentó el fracaso estructural del modelo de

aviso y consentimiento, que impone al individuo una carga de autogestión inasumible a escala; Edwards y Veale (2017) mostraron, en paralelo, la limitada eficacia del denominado derecho a la explicación frente a la opacidad técnica de los sistemas de aprendizaje automático. Desde esta perspectiva, la arquitectura de la Ley de IA resulta coherente: renuncia al consentimiento como técnica de tutela y lo sustituye por prohibiciones absolutas y obligaciones de diseño, de las que el artículo 5.1.f) constituye la manifestación más clara en el ámbito educativo.

De cuanto antecede se desprende una conclusión que matiza la lectura habitual sobre la complementariedad de ambos instrumentos. El Reglamento General de Protección de Datos y la Ley de IA no protegen la privacidad mediante una misma técnica jurídica: el primero conserva el consentimiento informado como pieza central de la licitud del tratamiento, mientras que la segunda opera al margen de él y, en el ámbito educativo, expresamente en contra de su virtualidad habilitante. Precisamente por ello, la remisión a las políticas públicas y a la autorregulación institucional no constituye un complemento accesorio del marco normativo, sino una exigencia derivada de su propia arquitectura: allí donde el consentimiento del estudiante no puede considerarse libre, la garantía debe trasladarse al diseño de la contratación, a la evaluación previa del sistema y a la decisión institucional sobre su despliegue.

En esta línea, Soengas (2025) recomienda a las IES revisar las políticas de privacidad de los proveedores para determinar si los datos se emplean en el entrenamiento de modelos o se comparten con terceros; priorizar herramientas conformes con el Reglamento General de Protección de Datos que ofrezcan garantías verificables sobre el tratamiento; comprobar si el proveedor está sujeto a la normativa de la Unión y adopta medidas anticipatorias respecto de la Ley de IA; y descartar aquellos servicios que no acrediten transparencia sobre el uso y el almacenamiento de datos en entornos educativos.

Estas cautelas resultan necesarias, pero insuficientes si no se acompañan de una fundamentación ética explícita. La cobertura normativa se ve desbordada tanto en el tiempo —por la distancia entre el ciclo legislativo y el ritmo de la innovación, que el propio aplazamiento operado por el Reglamento (UE) 2026/1744 ilustra— como en el alcance, pues ninguna norma puede anticipar la totalidad de los usos posibles ni la creciente demanda de datos que los sostiene. De ahí que los principios éticos que rigen el uso de la IA en la educación no operen como sustituto del derecho, sino como criterio de decisión institucional en el espacio que el derecho deja necesariamente abierto.

3. 5. El Consentimiento Informado en la Legislación Norteamericana

El enfoque de Estados Unidos se distingue, en general, por tener varias normas específicas en diferentes sectores y por los desarrollos en los estados, sin una ley federal completa que

se asemeje al RGPD. Esto ha sido señalado en la doctrina como un sistema regulatorio complicado y fragmentado (Barrio Andrés, 2022).

Este panorama normativo fragmentado refleja la necesidad de una aproximación variada y sectorial (servicios financieros, de salud, de telecomunicaciones y de educación) para la protección de los datos personales.

En el sector del comercio, en 2018, el estado de California aprobó la Ley de Privacidad del Consumidor de California (CCPA), la cual impuso por primera vez niveles de protección de datos muy similares a los establecidos en el RGPD. Posteriormente la CCPA fue modificada y ampliada por la Ley de Derechos de Privacidad de California (California Privacy Rights Act, CPRA).

En el caso de la educación, la normativa federal es la Ley de Derechos Educativos y Privacidad Familiar, mejor conocida por sus siglas FERPA, de 1974.

Para esta investigación nos centraremos en las dos leyes antes mencionadas, la FERPA y la CPRA.

3. 5. 1. Ley de Derechos de Privacidad de California (CPRA)

Una legislación relevante en el caso estadounidense es la Ley de Privacidad del Consumidor de California (California Consumer Privacy Act, CCPA), transformada posteriormente en la California Privacy Rights Act (CPRA)³. Este marco estatal busca reforzar derechos y obligaciones en materia de datos personales, incluidos deberes de transparencia y límites a prácticas manipulativas (California Secretary of State, 2020).

La CCPA fue un punto de inflexión en la creación de otros marcos regulatorios estatales en Estados Unidos. Esto se observa porque algunos estados han promulgado posteriormente sus propias leyes de privacidad⁴.

³ La CPRA es una ley de enmienda que modificó y amplió la CCPA (a menudo denominada en la práctica como «CCPA 2.0»). Aprobada mediante iniciativa popular (*Proposition 24*) en noviembre de 2020 y plenamente aplicable desde el 1 de enero de 2023.

⁴ A falta de una ley federal integral de privacidad en Estados Unidos (como lo intentó en su momento la *American Data Privacy and Protection Act* - ADPPA), múltiples estados promulgaron sus propias leyes generales de privacidad (*Comprehensive State Privacy Laws*) inspiradas directa o indirectamente en el marco de la CCPA y la CPRA. Virginia con la *Virginia Consumer Data Protection Act* (VCDPA, 2021); Colorado con la *Colorado Privacy Act* (CPA, 2021); Utah con la *Utah Consumer Privacy Act* (UCPA, 2022); Connecticut con la *Connecticut Data Privacy Act* (CTDPA, 2022); Iowa con la *Iowa Consumer Data Protection Act* (ICDPA, 2023); Indiana con la *Indiana Consumer Data Protection Act* (INCDPA, 2023); Tennessee con la *Tennessee Information Protection Act* (TIPA, 2023); Montana con la *Montana Consumer Data Privacy Act* (MTCDDPA, 2023); Texas con la *Texas Data Privacy and Security Act* (TDPSA, 2023); Oregon con la *Oregon Consumer Privacy Act* (OCPA, 2023); Delaware con la *Delaware Personal Data Privacy Act* (DPDPA, 2023); Nueva Jersey con la *New Jersey Data Privacy Act* (NJDDPA, 2024); Nuevo Hampshire con la *New Hampshire Privacy Act* (NHPA, 2024); Kentucky con la *Kentucky Consumer Data Privacy Act* (KCDPA, 2024); Nebraska

La CPRA comenzó a funcionar en 2023, pero la implementación de algunas reglas enfrentó problemas legales y se detuvo temporalmente. Una corte de apelación revisó esta situación en febrero de 2024 (Ropes & Gray, 2024; IAPP, 2024). La CPRA robusteció ese marco, aproximando el estándar norteamericano a los principios de gobernanza, minimización y supervisión propios del modelo europeo (GDPR).

La CPRA (2023), en la sección 1798. 140 contiene las definiciones de esta normativa y define de la siguiente manera la información personal:

(v. 1) “Personal information” means information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household. Personal information includes, but is not limited to, the following if it identifies, relates to, describes, is reasonably capable of being associated with, or could be reasonably linked, directly or indirectly, with a particular consumer or household⁵.

Por lo que se refiere al consentimiento, se hace una mención expresa de que este debe ser informado:

(h) “Consent” means any freely given, specific, informed, and unambiguous indication of the consumer's wishes by which the consumer, or the consumer's legal guardian, a person who has power of attorney, or a person acting as a conservator for the consumer, including by a statement or by a clear affirmative action, signifies agreement to the processing of personal information relating to the consumer for a narrowly defined particular purpose⁶.

con la *Nebraska Data Privacy Act* (NDPA, 2024); Maryland con la *Maryland Online Data Privacy Act* (MODPA, 2024); Minnesota con la *Minnesota Consumer Data Privacy Act* (MCDPA, 2024); y Rhode Island con la *Rhode Island Data Transparency and Privacy Protection Act* (RIDTPPA, 2024).

⁵ El texto traducido del inglés es el siguiente: "Información personal" significa información que identifica, se relaciona con, describe, es razonablemente capaz de ser asociada con, o podría razonablemente estar vinculada, directa o indirectamente, con un consumidor o hogar en particular. La información personal incluye, pero no se limita a, lo siguiente si identifica, se relaciona con, describe, es razonablemente capaz de ser asociada con, o podría ser razonablemente vinculada, directa o indirectamente, con un consumidor o hogar en particular. Traducción de la autora.

⁶ Por "consentimiento" se entenderá toda manifestación de voluntad libremente dada, específica, informada e inequívoca, mediante la cual el consumidor, su tutor legal, un apoderado o una persona que actúe como tutor del consumidor, incluso mediante una declaración o una clara acción afirmativa, manifieste su acuerdo con el tratamiento de información personal relativa al consumidor para un fin particular estrictamente definido. Traducción de la autora.

En esta misma legislación encontramos una declaración muy interesante desde el punto de vista de cómo obtener un auténtico consentimiento informado, ya que la normativa señala que el consentimiento no puede inferirse de la aceptación de unas condiciones generales o de uso, o de un documento similar, en el que se describan conjuntamente el tratamiento de datos personales y otra información no relacionada con el consentimiento. El texto detalla que no es consentimiento, pasar sobre un contenido, silenciarlo, pausarlo o cerrarlo (CPRA, 2023). Parece que con esta disposición se da a entender que debe haber un acto de voluntad activo y no una omisión.

Así, el texto de la CPRA incorpora una definición de “*dark patterns*” (patrones oscuros) relevante para evaluar la validez del consentimiento en interfaces digitales (*California Secretary of State*, 2020):

Acceptance of a general or broad terms of use, or similar document, that contains descriptions of personal information processing along with other, unrelated information, does not constitute consent. Hovering over, muting, pausing, or closing a given piece of content does not constitute consent. Likewise, agreement obtained through use of dark patterns does not constitute consent. (California Secretary of State, 2020)⁷.

A esta capa de regulaciones se añade el conjunto de normas que adoptó la *California Privacy Protection Agency* en 2025, que comienzan a aplicarse el 1 de enero de 2026. Estas nuevas normas incluyen obligaciones sobre evaluaciones de riesgo y el uso de tecnología de toma de decisiones automatizada (ADMT).

3. 5. 2. Ley de Derechos Educativos y Privacidad Familiar (FERPA)

La Ley de Derechos Educativos y Privacidad Familiar, conocida por sus siglas en inglés como FERPA, es también llamada Enmienda Buckley, en honor a James Buckley, senador de Estados Unidos que desempeñó un papel importante en su promulgación en 1974. Esta legislación federal fue introducida en respuesta a las crecientes preocupaciones respecto a la privacidad de los registros académicos de los estudiantes (FERPA, 1974). En ese momento, el interés sobre la privacidad y el acceso a la información personal en las instituciones educativas comenzó a ganar relevancia. La preocupación por el acceso no autorizado a datos personales y los abusos de información fueron factores clave que motivaron su creación (U. S. *Department of Education*, 2023a).

⁷ La aceptación de unos términos de uso generales o amplios, o de un documento similar, que contenga descripciones del tratamiento de datos personales junto con otra información no relacionada, no constituye consentimiento. Pasar el cursor por encima, silenciar, pausar o cerrar un contenido determinado no constituye consentimiento. Del mismo modo, el acuerdo obtenido mediante el uso de patrones oscuros no constituye consentimiento. Traducción de la autora.

El alcance de la ley es amplio porque es aplicable a todas las instituciones educativas que reciben fondos federales del Departamento de Educación de los Estados Unidos, abarcando tanto escuelas públicas como privadas y de educación superior, siempre que reciban financiamiento federal (U. S. *Department of Education*, 2023a).

La FERPA se creó con el objetivo de proteger la privacidad de los estudiantes, regular el acceso a la información confidencial y otorgar derechos específicos a los estudiantes y sus padres para controlar cómo se accede y se divulga dicha información.

En términos generales, la ley prohíbe que las instituciones divulguen información contenida en los registros académicos de los estudiantes sin obtener el consentimiento explícito del estudiante o, en su caso, de sus padres si el estudiante es menor de 18 años (FERPA, 1974).

Al detenerse a explicar con un poco más de detalle esta prohibición, se nota en sus elementos que la divulgación se refiere a cualquier acción que permita el acceso la revelación, la transferencia u otra forma de comunicación de información personal identificable contenida en registros educativos a cualquier parte, utilizando cualquier medio, ya sea verbal, escrito o electrónico (FERPA, 1974, actualización 10/31/2025)⁸.

En cuanto a los expedientes académicos, estos se definen de manera amplia e incluyen calificaciones, todos los cursos y la información personal sobre el estudiante, como información de contacto y fecha de graduación prevista (U. S. *Department of Education*, 2023b). Podemos concluir que se refieren a los datos relacionados con el desempeño académico de los estudiantes. En este contexto, la ley busca garantizar la privacidad de la información relacionada con el desempeño académico y el estatus de los estudiantes. Es importante destacar que el derecho de privacidad se transfiere de los padres al estudiante una vez que este cumple 18 años o al ingresar a una IES, momento en el cual se les considera "estudiantes elegibles" bajo la ley (FERPA, 1974, actualización 10/31/2025).

Además, la FERPA establece que los padres de los estudiantes menores de 18 años, o los propios estudiantes, si son mayores de edad o están matriculados en una IES, tienen el derecho de acceder a sus registros educativos, solicitar correcciones y controlar quienes pueden acceder a esa información. Esta disposición es similar a la que dan los derechos ARCO en México.

3. 5. 3. La Doctrina Especializada

La doctrina estadounidense ha cuestionado durante años la dependencia excesiva del modelo de notificación y elección. Tene y Polonetsky (2013) sostienen que el crecimiento del análisis masivo de datos exige combinar el control individual con limitación de finalidad,

⁸ Electronic Code of Federal Regulations (2024) 34 C.F.R. § 99.3 – *Definitions*. Disponible en: <https://www.ecfr.gov/current/title-34/subtitle-A/part-99>

minimización y responsabilidad institucional. La capacidad de inferir nueva información dificulta que una persona anticipe todos los usos al momento de consentir.

Solove (2013) denomina a este problema el “dilema del consentimiento”. La autogestión de la privacidad supone que las personas pueden leer, comprender y evaluar las consecuencias de cada política, pero la complejidad, la cantidad de decisiones y los efectos acumulativos hacen que ese control sea frecuentemente ilusorio. El problema no se resuelve simplemente proporcionando avisos más extensos. Reidenberg y Schaub (2018) enfatizan la necesidad de un consentimiento informado, asegurando que los estudiantes y sus tutores comprendan qué datos se recopilan, cómo se utilizan y con quién se comparten. Este enfoque busca reducir la asimetría de información y otorgar a los estudiantes un mayor control sobre su privacidad en un entorno digitalizado.

Algunos académicos argumentan que la legislación actual, incluida la FERPA, es insuficiente para abordar los desafíos emergentes en materia de privacidad de datos en el ámbito educativo (Reidenberg y Schaub, 2018). La ley fue diseñada para proteger registros académicos tradicionales y no aborda la gran cantidad de datos generados hoy en día por sistemas digitales en la educación (Borgman, 2018).

El aumento masivo de datos recopilados por las instituciones educativas a través de plataformas de aprendizaje en línea, sistemas de gestión educativa, y analíticas de aprendizaje está generando nuevos desafíos para la privacidad de los estudiantes, debido a que estos datos incluyen interacciones en tiempo real, resultados académicos detallados y patrones de comportamiento estudiantil, lo que complica su inclusión dentro de la categoría tradicional de educational records o registros escolares contemplados por la FERPA (*Public Interest Privacy*, 2024).

Zeide (2016) sostiene que FERPA fue diseñada para gobernar instituciones educativas financiadas federalmente y no para regular directamente ecosistemas comerciales complejos. La autora cuestiona que la protección pueda depender de la autogestión individual y de una supervisión institucional meramente teórica. En entornos de big data, la protección requiere responsabilidades sustanciales para las instituciones y los receptores de información, no solo autorizaciones formales. Por lo anterior, algunos académicos abogan por la implementación de un marco normativo más robusto. Este marco debe establecer con claridad la propiedad de los datos y garantizar que los estudiantes tengan derechos explícitos sobre la información que generan a través de plataformas digitales y herramientas de aprendizaje (Reidenberg y Schaub, 2018; Borgman, 2018).

Aunado a lo anterior, está el hecho de que las leyes estatales que protegen la privacidad varían de estado en estado; a esta dispersión normativa en EE. UU. se le llama el mosaico (*patchwork*) de privacidad estadounidense.

Dado que una universidad o plataforma EdTech puede procesar datos de estudiantes residentes en distintos estados de EE. UU. o de la Unión Europea, resulta inviable operativamente diseñar un sistema de cumplimiento individualizado para cada ley local. Por ello, la implementación de un sistema de gestión como ISO IEC 27701 (PIMS) o ISO IEC 42001 (AIMS) funciona como un "paraguas de cumplimiento normativo (*compliance*) neutral", permitiendo a la institución cumplir de forma simultánea con la CPRA, las demás leyes estatales de EE. UU. y el RGPD europeo a través de una sola arquitectura de gobernanza.

3. 6. El Consentimiento Informado en México

El artículo 6º de la Constitución Política de los Estados Unidos Mexicanos consagra el derecho de acceso a la información pública, mientras que el artículo 16, párrafo segundo —modificado en diversas ocasiones— reconoce de manera autónoma el derecho de toda persona a la protección de sus datos personales, así como al acceso, rectificación y cancelación de los mismos y a manifestar su oposición al tratamiento. Se trata de dos derechos fundamentales distintos, con sedes constitucionales diferenciadas, cuya confusión resulta frecuente en la literatura.

El 20 de diciembre de 2024 se publicó en el Diario Oficial de la Federación (DOF) el “Decreto por el que se reforman, adicionan y derogan diversas disposiciones de la Constitución Política de los Estados Unidos Mexicanos, en materia de simplificación orgánica”, cuyo objeto principal consistió en extinguir diversos organismos constitucionales autónomos, entre ellos el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI).

En desarrollo de esa reforma, el DOF del 20 de marzo de 2025 publicó el decreto por el que se expiden la Ley General de Transparencia y Acceso a la Información Pública, la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO) y la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), y se reforma el artículo 37, fracción XV, de la Ley Orgánica de la Administración Pública Federal. Ambas leyes en materia de datos personales entraron en vigor el 21 de marzo de 2025 y abrogaron, respectivamente, la Ley General publicada el 26 de enero de 2017 y la Ley Federal publicada el 5 de julio de 2010 (Diario Oficial de la Federación, 20 de marzo de 2025, transitorio segundo).

Con posterioridad, el DOF del 14 de noviembre de 2025 publicó el “Decreto por el que se reforman diversas disposiciones de diversos ordenamientos legales, en materia de homologación normativa relativa al Código Nacional de Procedimientos Civiles y Familiares”, cuyo artículo trigésimo octavo reformó el artículo 4 de la LFPDPPP y que alcanzó igualmente al artículo 9 de la LGPDPPSO. Conviene precisar que la aplicación de dicho decreto no es inmediata, sino gradual, conforme a las declaratorias que expidan los

congresos locales y el Congreso de la Unión para la entrada en vigor del Código Nacional de Procedimientos Civiles y Familiares, y que operará de forma automática a partir del 1º de abril de 2027 en caso de no emitirse. Para efectos de esta tesis, el marco mexicano se cita conforme a su texto vigente con corte a esa reforma, identificando en cada caso el decreto y la fecha de publicación, a fin de evitar referencias genéricas o no trazables (Diario Oficial de la Federación, 14 de noviembre de 2025).

3. 6. 1. Ley Federal de Protección de Datos Personales en Posesión de los Particulares

La ley de 2025 introduce modificaciones sustantivas respecto de la abrogada. Entre ellas destacan la ampliación de la definición de tratamiento, que comprende cualquier operación efectuada mediante procedimientos manuales o automatizados (LFPDPPP, 2025, art. 2, fracc. XIX); el reforzamiento del deber de confidencialidad; y la incorporación del derecho de oposición frente al tratamiento automatizado (Téllez Domínguez, 2025). Se amplía asimismo el universo de obligados, pues el artículo 2, fracciones XIV y XVI, define al responsable por remisión a la noción de “sujetos regulados”, entendidos como las personas físicas o morales de carácter privado que llevan a cabo el tratamiento de datos personales, sin exigir que adopten decisiones sobre él.

El derecho de oposición frente al tratamiento automatizado merece precisión, dada su relevancia para esta investigación. El artículo 26, fracción II, faculta a la persona titular para oponerse cuando sus datos “sean objeto de un tratamiento automatizado, el cual le produzca efectos jurídicos no deseados o afecte de manera significativa sus intereses, derechos o libertades, y estén destinados a evaluar, sin intervención humana, determinados aspectos personales de la misma o analizar o predecir, en particular, su rendimiento profesional, situación económica, estado de salud, preferencias sexuales, fiabilidad o comportamiento”. El mismo precepto excluye el ejercicio del derecho cuando el tratamiento resulte necesario para el cumplimiento de una obligación legal impuesta al responsable. Debe subrayarse que la ley no emplea en ningún punto de su articulado la expresión “IA”: el supuesto normativo se construye sobre la noción de tratamiento automatizado sin intervención humana, lo que comprende los sistemas de IA sin mencionarlos, y sin establecer para ellos régimen específico alguno.

De la legislación anterior se conserva la obligación del responsable de observar los principios de licitud, finalidad, lealtad, consentimiento, calidad, proporcionalidad, información y responsabilidad en el tratamiento de datos personales (LFPDPPP, 2025, art. 5).

De particular relevancia para este campo de investigación resulta la centralidad que la reforma otorga al consentimiento. El artículo 2, fracción IV, lo define como la “manifestación de la voluntad libre, específica e informada de la persona titular de los datos mediante la cual se efectúa el tratamiento de los mismos”. Por su parte, el artículo 11 dispone que el tratamiento deberá limitarse al cumplimiento de las finalidades previstas en el aviso de

privacidad y que, si el responsable pretende tratar los datos para una finalidad distinta de las establecidas en él, se requerirá obtener nuevamente el consentimiento de la persona titular. Esta redacción suprime la ambigüedad que permitía el régimen anterior, cuyo artículo 12 exigía nuevo consentimiento únicamente cuando la finalidad distinta “no resulte compatible o análoga” a las previstas en el aviso (LFPDPPP, 2010, art. 12). La eliminación de esa salvedad estrecha de manera apreciable el margen de reutilización de los datos. Algunos autores consideran que la reforma busca contrarrestar las prácticas habituales de las plataformas digitales, en las que el consentimiento resulta genérico, poco claro o forzoso (Téllez Domínguez, 2025).

Conviene precisar que, aunque podría sostenerse que la protección de la privacidad no resulta del todo efectiva en la medida en que el artículo 7 reconoce como regla general la validez del consentimiento tácito, dicha previsión no opera en términos absolutos, toda vez que el artículo 8 exige, tratándose de datos personales sensibles, el consentimiento expreso y por escrito de la persona titular, mediante firma autógrafa, firma electrónica o cualquier mecanismo de autenticación que al efecto se establezca. Ello sin perjuicio de las excepciones al consentimiento previstas con carácter general en el artículo 9 (LFPDPPP, 2025, arts. 7-9).

En materia de seguridad y tutela de derechos, la nueva regulación obliga al responsable o tercero a establecer controles o mecanismos que aseguren que todas las personas que intervengan en cualquier fase del tratamiento guarden confidencialidad respecto de los datos, obligación que subsiste aun después de finalizadas sus relaciones con el responsable (LFPDPPP, 2025, art. 20). En cuanto a las transferencias, el aviso de privacidad debe contener una cláusula en la que se indique si la persona titular acepta o no la transferencia de sus datos, salvo que se actualice alguno de los supuestos legalmente exceptuados (LFPDPPP, 2025, arts. 35-36). En congruencia con ello, el artículo 58, fracción XIII, tipifica como infracción “recabar o transferir datos personales sin el consentimiento expreso de la persona titular, en los casos en que éste sea exigible”.

La noción legal de consentimiento informado se articula con la de aviso de privacidad, definido en el artículo 2, fracción I, como el “documento a disposición de la persona titular de la información de forma física, electrónica o en cualquier otro formato generado por el responsable, a partir del momento en el cual se recaben sus datos personales, con el objeto de informarle los propósitos del tratamiento de los mismos, de conformidad con el artículo 14 de la presente Ley”.

El aviso de privacidad sigue siendo, por tanto, el instrumento clave a través del cual se cumple el principio de información (Quijano Decanini, 2022). La ley atiende además a su presentación en entornos digitales: el artículo 16, fracción II, dispone que, cuando los datos personales se obtengan por medios electrónicos, ópticos, sonoros, visuales o cualquier otra tecnología, el aviso deberá proporcionarse en modalidad simplificada, con el contenido

mínimo de las fracciones I a IV del artículo 15 y la indicación del sitio donde pueda consultarse el aviso integral.

3. 6. 2. Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados

Esta legislación define el consentimiento en términos idénticos a los de la Ley Federal (LGPDPPO, 2025, art. 3, fracc. VIII) y recoge los principios de licitud, finalidad, lealtad, consentimiento, calidad, proporcionalidad, información y responsabilidad en su artículo 10. Entre las causales que eximen al responsable de recabar el consentimiento de la persona titular, el artículo 16 contempla el supuesto de que los datos personales se requieran para ejercer un derecho o cumplir obligaciones derivadas de una relación jurídica entre la persona titular y el responsable (LGPDPPO, 2025, art. 16, fracc. V).

Las universidades públicas autónomas son sujetos obligados por la LGPDPPSO. En consecuencia, no están obligadas a recabar el consentimiento del titular —mediante el aviso de privacidad— cuando los datos resulten necesarios para la prestación del servicio educativo derivado de la relación de matrícula. La excepción cubre el tratamiento indispensable para ese fin, y no cualquier tratamiento que traiga causa de dicha relación.

3. 6. 3. El Consentimiento de los Estudiantes: Comparación entre IES Públicas y Privadas

El uso de IA y de otras tecnologías digitales en las IES implica la recolección, conservación, análisis y en su caso, la transferencia de datos personales del estudiantado.

En el ordenamiento jurídico mexicano coexisten dos regímenes aplicables: la LFPDPPP, para las instituciones particulares, y la LGPDPPSO, para las pertenecientes al sector público. Ambas legislaciones presentan similitudes relevantes, pero también diferencias que resulta necesario considerar y que se desarrollan en los subapartados siguientes.

3. 6. 3. 1. Concepto Normativo de Consentimiento y sus Formas de Expresión

Ambas leyes definen el consentimiento como manifestación de la voluntad libre, específica e informada (LFPDPPP, 2025, art. 2, fracc. IV; LGPDPPSO, 2025, art. 3, fracc. VIII). Sin embargo, las lógicas de licitud son distintas: la LFPDPPP establece el consentimiento como regla general del tratamiento, mientras que la LGPDPPSO lo condiciona, desde el inicio, a las facultades y atribuciones normativas del sujeto obligado.

En el caso de las instituciones privadas, el consentimiento opera como condición habilitante primaria del tratamiento. En las públicas, en cambio, se integra en un marco de legalidad administrativa, conforme al cual el tratamiento debe sujetarse a las facultades o atribuciones expresamente previstas en la normatividad aplicable (LGPDPPO, 2025, art. 11).

En ambos regímenes el consentimiento puede otorgarse de forma expresa o tácita. Es expresa cuando la voluntad se manifiesta verbalmente, por escrito, por medios electrónicos, ópticos, signos inequívocos o cualquier otra tecnología; es tácita cuando, habiéndose puesto a disposición el aviso de privacidad, la persona titular no manifiesta su voluntad en sentido contrario (LFPDPPP, 2025, art. 7; LGPDPPSO, 2025, art. 15).

Uno y otro ordenamiento establecen reglas reforzadas para determinadas categorías de datos, si bien con alcance distinto. La LFPDPPP exige consentimiento expreso para los datos financieros o patrimoniales, salvo las excepciones de sus artículos 9 y 36, y consentimiento expreso y por escrito para los datos personales sensibles, mediante firma autógrafa, firma electrónica o mecanismo de autenticación equivalente (LFPDPPP, 2025, arts. 7-8). La LGPDPPSO no contiene mención expresa a los datos financieros o patrimoniales, pero impone idéntica exigencia de consentimiento expreso y por escrito respecto de los datos personales sensibles, salvo en los casos previstos en su artículo 16 (LGPDPPSO, 2025, art. 15). El estándar reforzado aplicable a los datos sensibles es, por tanto, común a ambos regímenes.

3. 6. 3. 2. Aviso de Privacidad, el Instrumento Clave del Consentimiento Informado

El aviso de privacidad actúa como puente entre los principios de información y consentimiento. El responsable tiene la obligación de informar a la persona titular, a través de dicho documento, la existencia y características principales del tratamiento al que serán sometidos sus datos, a fin de que pueda tomar decisiones informadas al respecto (LFPDPPP, 2025, art. 14).

En el régimen aplicable a los particulares, el aviso debe contener, como mínimo, la identidad y domicilio del responsable; los datos personales que serán sometidos a tratamiento, identificando aquellos que son sensibles; las finalidades del tratamiento, distinguiendo las que requieren el consentimiento de la persona titular; las opciones y medios para limitar el uso o divulgación de los datos; los mecanismos, medios y procedimientos para ejercer los derechos ARCO —acceso, rectificación, cancelación y oposición—; y el procedimiento y medio por el cual se comunicarán los cambios al propio aviso (LFPDPPP, 2025, art. 15). Bajo este marco, la incorporación de sistemas de IA en procesos universitarios —mecanismos de admisión, herramientas de supervisión remota de exámenes o tutorías inteligentes— exige que el aviso de privacidad delimite con claridad las finalidades esenciales y accesorias del tratamiento e identifique explícitamente aquellas que requieren consentimiento, a fin de evitar tratamientos implícitos o normativamente ambiguos.

En el sector privado, además, el tratamiento debe limitarse a las finalidades previstas en el aviso, de modo que si el responsable pretende tratar los datos para una finalidad distinta, se requerirá obtener nuevamente el consentimiento (LFPDPPP, 2025, art. 11).

En el caso de los sujetos obligados, se reitera la centralidad del aviso de privacidad y se exige que incluya elementos análogos a los del régimen privado, añadiendo de forma expresa el fundamento jurídico que faculta el tratamiento, así como los mecanismos para que la persona titular conozca el aviso integral cuando se emplee la modalidad simplificada (LGPDPPO, 2025, arts. 21-22).

El tratamiento en el sector público se enmarca, además, en una premisa adicional: debe sujetarse a las facultades o atribuciones conferidas por la normatividad aplicable (LGPDPPO, 2025, art. 11) y estar justificado por finalidades concretas, lícitas, explícitas y legítimas relacionadas con dichas atribuciones (LGPDPPO, 2025, art. 12). De ello se sigue que, si una institución pública pretende reutilizar datos recabados para una finalidad institucional —por ejemplo, la gestión escolar— con fines distintos; por ejemplo, el entrenamiento de modelos predictivos, no basta con incorporar una cláusula de consentimiento: debe acreditarse también la atribución normativa que legitima ese nuevo tratamiento, además de recabar el consentimiento cuando corresponda (LGPDPPO, 2025, art. 12).

Para valorar la fuerza vinculante de la figura del consentimiento informado resulta oportuno examinar las consecuencias jurídicas de la omisión o deficiencia del aviso de privacidad.

Tratándose de una institución privada, omitir en el aviso alguno o todos los elementos exigidos por el artículo 15 constituye la infracción prevista en el artículo 58, fracción V, sancionada con multa de cien a ciento sesenta mil veces la Unidad de Medida y Actualización (LFPDPPP, 2025, art. 59, fracc. II). Las conductas de mayor gravedad —entre ellas llevar a cabo transferencias fuera de los casos permitidos por la ley, o recabar y transferir datos sin el consentimiento expreso exigible (art. 58, fracciones XII y XIII)— se sancionan con multa de doscientas a trescientas veinte mil veces dicha unidad (art. 59, fracc. III). En caso de reincidencia procede una multa adicional de cien a trescientas veinte mil veces la misma unidad y, tratándose de infracciones cometidas en el tratamiento de datos personales sensibles, las sanciones podrán incrementarse hasta por dos veces los montos establecidos (art. 59, fracc. IV). A ello se añade el régimen penal previsto en los artículos 62 a 64, que sanciona con prisión de tres meses a tres años a quien, estando autorizado para tratar datos personales y con ánimo de lucro, provoque una vulneración de seguridad, y con prisión de seis meses a cinco años a quien, con el fin de alcanzar un lucro indebido, trate datos personales mediante engaño; penas que se duplican tratándose de datos sensibles.

En las instituciones públicas, en cambio, el incumplimiento de obligaciones tales como la ausencia del aviso de privacidad o la omisión de los elementos legalmente exigidos se tipifica como causa de sanción y se vincula al régimen de responsabilidades administrativas aplicable a las personas servidoras públicas (LGPDPPO, 2025, art. 132).

Más allá del aviso de privacidad es importante destacar que la Encuesta Nacional de IA Generativa (ENIAG) aporta evidencia empírica directa sobre la brecha entre la existencia formal de normativa -el consentimiento- y su conocimiento efectivo por parte de la comunidad universitaria: más del 76% de los estudiantes y el 75% de los docentes de instituciones públicas declaran no conocer ninguna normativa institucional sobre el uso de la IA (SEP, 2025). La misma brecha, aunque algo menor, se observa en instituciones particulares, donde el desconocimiento se ubica entre el 61% y el 67% según el grupo consultado. Por lo tanto, la existencia de un aviso de privacidad o de un documento normativo no garantiza que el consentimiento informado opere realmente como tal, pues el estudiantado y el profesorado desconocen, en su mayoría, las reglas bajo las cuales sus datos son tratados.

3. 6. 3. 3. Excepciones a la Obtención del Consentimiento y su Impacto en el Diseño Institucional

Ambos regímenes contemplan causales que permiten al responsable tratar datos sin recabar el consentimiento de la persona titular, con el objeto de armonizar la protección de los datos con otros bienes jurídicos.

La LFPDPPP exonera de recabar el consentimiento cuando una disposición jurídica así lo disponga; cuando los datos figuren en fuentes de acceso público; cuando se sometan a un procedimiento previo de disociación; cuando se requieran para ejercer un derecho o cumplir obligaciones derivadas de una relación jurídica entre la persona titular y el responsable; ante situaciones de emergencia que puedan dañar a una persona o sus bienes; cuando sean indispensables para la atención médica mientras la persona titular no esté en condiciones de otorgar el consentimiento; y cuando exista orden judicial, resolución o mandato fundado y motivado de autoridad competente (LFPDPPP, 2025, art. 9).

La LGPDPPSO establece excepciones análogas —disposición legal, fuentes de acceso público, disociación y finalidades estadísticas o científicas cuando los datos no puedan asociarse a la persona titular—, y añade supuestos ligados a la operatividad de las atribuciones institucionales, lo que refuerza la lógica de la legalidad administrativa (LGPDPPSO, 2025, art. 16).

Para facilitar la lectura del análisis, la Tabla 1 presenta una síntesis de las diferencias relevantes entre ambos regímenes.

Tabla 1

Eje	IES privadas (LFPDPPP)	IES públicas / sujetos obligados (LGPDPPO)
Base de licitud (regla general)	Tratamiento sujeto al consentimiento salvo excepciones (art. 7).	Tratamiento sujeto a atribuciones y finalidades vinculadas a la normatividad aplicable (arts. 11–12).
Concepto de consentimiento	Manifestación libre, específica e informada (art. 2, fracc. IV).	Manifestación libre, específica e informada (art. 3, fracc. VIII).
Expreso / tácito; sensibles	Expreso o tácito; financieros/patrimoniales requieren expreso; sensibles requieren expreso y por escrito (arts. 7–8).	Expreso o tácito (art. 15) y criterios de libertad/especificidad/información (art. 14).
Aviso de privacidad	Contenido mínimo detallado (art. 15) y puesta a disposición por múltiples medios/tecnologías (art. 16).	Debe incluir, entre otros, fundamento jurídico del tratamiento y mecanismos para conocer aviso integral (arts. 21–22).
Cambio de finalidad	Nueva finalidad exige nuevo consentimiento (art. 11).	Finalidad distinta exige atribución normativa y consentimiento (art. 12).
Riesgo sancionatorio	Infracción por omitir aviso (art. 58); multas 100–320, 000 UMA (art. 59) y delitos (art. 63).	Medidas de apremio (art. 122) y causas de sanción ligadas a responsabilidades administrativas (art. 132).
<i>Fuente: Elaboración propia</i>		

Del contraste entre la LFPDPPP y la LGPDPPSO cabe concluir que el consentimiento, aunque central en ambos regímenes, no posee la misma capacidad legitimante: en el sector privado opera como regla de habilitación del tratamiento, mientras que en el sector público se integra en un marco de atribuciones y finalidades que limita su alcance.

3. 6. 4. Transferencias, Remisiones y Contratación de Proveedores de IA

El ordenamiento mexicano distingue dos categorías de comunicación de datos personales con regímenes jurídicos radicalmente distintos. La LGPDPPSO define la remisión como toda comunicación realizada exclusivamente entre el responsable y la persona encargada, dentro o fuera del territorio mexicano, y la transferencia como toda comunicación realizada a persona distinta de la titular, del responsable o de la persona encargada (LGPDPPO, 2025, art. 3, fracciones XXIV y XXX). La LFPDPPP no define la remisión, pero adopta el mismo criterio al excluir expresamente a la persona encargada del concepto de transferencia (LFPDPPP, 2025, art. 2, fracc. XX). El deslinde no atiende a la identidad ni a la ubicación del receptor, sino a la titularidad del poder de decisión sobre el tratamiento: la persona encargada es aquella, ajena a la organización del responsable, que trata datos a nombre y por cuenta de este (LGPDPPO, 2025, art. 3, fracc. XXII; LFPDPPP, 2025, art. 2, fracc. XII), y

debe hacerlo sin ostentar poder alguno de decisión sobre el alcance y contenido del tratamiento, limitando sus actuaciones a los términos fijados por el responsable (LGPDPPO, 2025, art. 52). La calificación es, por tanto, funcional y no nominal: no depende de la denominación contractual, sino de si el proveedor conserva capacidad de decisión sobre las finalidades.

La consecuencia de la calificación es determinante. Las remisiones nacionales e internacionales entre el responsable y la persona encargada no requieren ser informadas a la persona titular ni contar con su consentimiento (LGPDPPO, 2025, art. 65), pues jurídicamente no hay comunicación a un tercero: el encargado opera como prolongación instrumental del responsable, quien conserva íntegramente su posición frente al titular. La exención de consentimiento no equivale, sin embargo, a la ausencia de garantías. La relación debe formalizarse mediante contrato o instrumento jurídico equivalente que acredite su existencia, alcance y contenido, y que prevea al menos siete cláusulas: tratar los datos conforme a las instrucciones recibidas, abstenerse de destinarlos a finalidades distintas de las instruidas, implementar medidas de seguridad, informar las vulneraciones, guardar confidencialidad, suprimir o devolver los datos al concluir la relación y abstenerse de transferirlos salvo determinación del responsable, subcontratación autorizada o mandato de autoridad (LGPDPPO, 2025, art. 53).

Las remisiones fuera del territorio nacional exigen además que la persona encargada se obligue a proteger los datos conforme a los principios y deberes de la LGPDPO (art. 62), y la subcontratación requiere autorización expresa, adquiriendo la persona subcontratada el carácter de encargada (arts. 55 y 56).

Toda transferencia, nacional o internacional, se encuentra en cambio sujeta al consentimiento de la persona titular, salvo las excepciones de los artículos 16, 60 y 64 de la propia LGPDPO (art. 59), y debe formalizarse mediante cláusulas contractuales, convenios de colaboración u otro instrumento jurídico que demuestre el alcance del tratamiento y las obligaciones asumidas por las partes (art. 60), además de comunicar al receptor el aviso de privacidad correspondiente (art. 63).

Aquí emerge una asimetría de la que depende el carácter informado del consentimiento: en el régimen público, el aviso debe identificar a quién se transfieren los datos y con qué finalidades, y prever los medios para que la persona titular manifieste su negativa antes de que el tratamiento ocurra (LGPDPPO, 2025, art. 21, fracciones VII y VIII); en el régimen privado, el artículo 15 de la LFPDPPP suprimió la obligación de informar sobre las transferencias que contemplaba la legislación abrogada, conservando únicamente la exigencia de incluir en el aviso una cláusula de aceptación o rechazo (LFPDPPP, 2025, art. 35, párr. segundo). El estudiantado de una institución privada conserva así la facultad de aceptar o negarse, pero no el derecho a conocer el destinatario.

Para las IES, la cuestión decisiva consiste en determinar en cuál de ambas categorías se ubica el proveedor de un sistema de IA, y la respuesta depende de una sola variable: si se limita a procesar los insumos conforme a las instrucciones institucionales, o si los destina a finalidades propias, señaladamente el entrenamiento o la mejora de sus modelos. El ordenamiento anticipa expresamente el segundo supuesto y le asigna una consecuencia precisa: cuando la persona encargada incumple las instrucciones del responsable y decide por sí misma sobre el tratamiento, asume el carácter de responsable y las consecuencias legales correspondientes (LGPDPPO, 2025, art. 54). No deviene, pues, un simple tercero receptor: adquiere por ministerio de ley la condición de responsable, con la totalidad de los deberes que ello comporta —base de licitud propia, aviso de privacidad, medidas de seguridad y atención de derechos ARCO—, en tanto que la comunicación deja de ser remisión y queda sujeta al régimen de la transferencia.

De lo anterior se sigue que el momento jurídicamente decisivo no es aquel en que la institución recaba el consentimiento de su comunidad, sino el anterior, en que examina las condiciones del servicio y suscribe el instrumento jurídico correspondiente: es en la cláusula relativa al uso de los datos de entrada donde se determina la calidad con que actuará el proveedor y, con ella, todo el régimen aplicable. La LGPDPPSO refuerza esta conclusión al regular específicamente la contratación por adhesión de servicios de cómputo en la nube, permitida sólo cuando la persona proveedora tenga políticas de protección afines a los principios y deberes legales, transparente sus subcontrataciones, se abstenga de incluir condiciones que le autoricen a asumir la titularidad o propiedad de la información, guarde confidencialidad, permita al responsable limitar el tipo de tratamiento y garantice la supresión de los datos al concluir el servicio; y establece que en ningún caso podrá el responsable adherirse a servicios que no garanticen la debida protección de los datos personales (LGPDPPO, 2025, arts. 57 y 58). El consentimiento posterior del estudiantado no subsana una adhesión contraria a este precepto, cuya prohibición no se condiciona a la voluntad del titular sino a una cualidad objetiva del servicio.

Cabe cerrar con una constatación que confirma la tesis sostenida en este capítulo. La LGPDPPSO destina un título completo a la relación entre responsable y persona encargada (arts. 52 a 58) y otro a las comunicaciones de datos personales (arts. 59 a 65), regulando la formalización contractual, el contenido mínimo de las cláusulas, la subcontratación, el cambio de calificación jurídica y la contratación de servicios en la nube. La LFPDPPP carece de todo ello: menciona a la persona encargada únicamente en el catálogo de definiciones, en la delimitación negativa de la transferencia y a efectos de responsabilidad civil, sin imponer deber alguno de formalización, sin regular la subcontratación y sin norma equivalente sobre cómputo en la nube, materia que la legislación abrogada remite a su Reglamento y que hoy carece de desarrollo.

La consecuencia es que una universidad pública que contrata un sistema de IA queda sujeta a un régimen legal detallado y exigible, mientras que una universidad privada que contrata

el mismo sistema opera en un vacío normativo de rango legal. La asimetría no beneficia al régimen aparentemente más flexible: lo priva de las herramientas jurídicas para exigir garantías a sus proveedores y traslada íntegramente a la autorregulación institucional una función que en el sector público desempeña la ley.

3. 6. 5. La Reconfiguración de las Autoridades Garantes

El análisis de los regímenes de consentimiento no sería completo sin considerar a la entidad que supervisa su cumplimiento. La efectividad de una garantía no solo depende de cómo está escrita la norma, sino también de tener un órgano que pueda hacerla cumplir. En este plano, la reforma mexicana de 2024-2025 introdujo una transformación de mayor calado que las modificaciones sustantivas examinadas hasta aquí.

El Decreto de reforma constitucional en materia de simplificación orgánica, publicado el 20 de diciembre de 2024, extinguió el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales. Este órgano constitucional autónomo había concentrado, desde su creación, las funciones de supervisión, verificación y sanción en la materia respecto de ambos sectores. Su desaparición no fue acompañada de la creación de un órgano equivalente, sino de una redistribución de competencias que opera de manera distinta en cada régimen.

En el ámbito de los particulares, las atribuciones del extinto Instituto se transfirieron a la Secretaría Anticorrupción y Buen Gobierno, dependencia del Ejecutivo Federal. La supervisión cambia de un organismo con autonomía constitucional a una parte de la administración centralizada, lo que provoca una pérdida de independencia frente al poder político. A ello se añade que el reglamento de la nueva Ley Federal, cuya expedición debía producirse dentro de los noventa días naturales siguientes a su entrada en vigor, permanece pendiente, de modo que las reglas de desarrollo del aviso de privacidad y de las transferencias continúan rigiéndose supletoriamente por el reglamento de 2011 en lo que no contradiga la ley vigente.

En el ámbito de los sujetos obligados, la solución adoptada es de naturaleza distinta y de consecuencias más profundas: la función de autoridad garante en materia de protección de datos personales corresponde al órgano interno de control de cada sujeto obligado. Aplicada a una universidad pública autónoma, la regla significa que la instancia encargada de verificar la licitud del tratamiento de los datos de su comunidad, de conocer de las inconformidades y de dar vista para la imposición de sanciones (LGPDPSO, 2025, art. 132) forma parte de la propia estructura institucional cuya actuación debe supervisar. La garantía deja de ser externa y se internaliza.

De la conjunción de ambos elementos resulta una paradoja que conviene enunciar con precisión. Como se ha mostrado en los apartados precedentes, el régimen sustantivo aplicable

a las instituciones públicas es materialmente más exigente que el aplicable a las privadas: el tratamiento debe sujetarse a las facultades o atribuciones conferidas por la normatividad aplicable (LGPDPPO, 2025, art. 11) y toda finalidad distinta de las previstas en el aviso de privacidad requiere, de manera acumulativa, atribución normativa y consentimiento de la persona titular (LGPDPPO, 2025, art. 12). Sin embargo, ese régimen más estricto queda confiado a una supervisión interna, mientras que el régimen más permisivo de las instituciones privadas conserva, al menos formalmente, una autoridad externa. La intensidad de la exigencia sustantiva y la de la garantía institucional operan, por tanto, en sentido inverso.

El contraste con el ordenamiento europeo, analizado en el apartado 3.4, resulta ilustrativo. Allí, el debate suscitado por el paquete Digital Omnibus durante 2025 y 2026 versó sobre si determinadas enmiendas reducían el nivel de protección vigente, y fueron precisamente las autoridades independientes de protección de datos —el Comité Europeo de Protección de Datos y el Supervisor Europeo— quienes formularon la advertencia (EDPB & EDPS, 2026). La existencia de un órgano externo e independiente constituye, en ese modelo, la condición de posibilidad de la crítica institucional. En el caso mexicano, durante el mismo periodo, la autoridad independiente fue suprimida y, tratándose de sujetos obligados, la función garante se internalizó. La divergencia no afecta al contenido de los derechos reconocidos, que es sustancialmente análogo, sino a la arquitectura destinada a hacerlos exigibles.

De cuanto antecede se desprende la conclusión que articula este capítulo. La protección de los datos personales del estudiantado frente a los sistemas de IA descansa, en el ordenamiento mexicano, sobre una cadena de garantías cuyos eslabones presentan debilidades sucesivas: el consentimiento no puede reputarse libre en una relación estructuralmente asimétrica; el aviso de privacidad integral ya no debe informar sobre las transferencias que realice el responsable; las reglas reglamentarias de desarrollo están pendientes de expedición; y la supervisión, en el caso de las instituciones públicas, corresponde a un órgano de la propia institución. En tales condiciones, los marcos institucionales de uso ético de la IA no operan como complemento voluntario de la normativa, sino como la principal garantía efectivamente disponible. Es en esa función —y no en su valor declarativo— donde radica su relevancia jurídica, y es lo que justifica el análisis comparado que se desarrolla en el capítulo siguiente.

3. 6. 6. La Ley General de la Educación

La actual ley es del 2019 y su última reforma publicada es del 15-01-2026. El artículo 170, fracción XXV, de dicho ordenamiento es el único precepto que vincula datos y consentimiento, y lo hace mediante una fórmula ambigua: "su titular o, en su caso, de la madre y padre de familia o tutor". La construcción sugiere que la titularidad puede corresponder al propio educando, pero desplaza la manifestación de voluntad hacia quienes ejercen la patria potestad sin establecer criterios de edad, madurez o autonomía progresiva. Este patrón se confirma en el resto de la Ley: el consentimiento informado exigido para

administrar medicamentos (art. 170, fracción XIII) es de madres, padres o tutores; el catálogo de derechos del educando (art. 72) no incluye la autodeterminación informativa ni el acceso, rectificación, cancelación u oposición; y los artículos 21, 128 (fracciones VI a X) y 129 configuran, más bien, un flujo institucional de información hacia las familias sobre desempeño y conducta del estudiante. Estos artículos no aplican a las IES cuando el estudiante es mayor de edad y por lo tanto es el titular de los derechos de privacidad.

La única figura donde el educando concurre a la decisión es la valoración previa para la prestación de educación especial (arts. 7, fracción II, inciso d, y 64, fracción I), y ahí se trata de una decisión pedagógica compartida, no de un consentimiento sobre datos.

En consecuencia, cualquier marco de gobernanza de datos o de uso ético de IA en el ámbito educativo mexicano para las IES debe apoyarse en la legislación general de protección de datos, antes expuestas, y en las normativas institucionales propia, pues la Ley General de Educación resulta insuficiente y, en materia de consentimiento estudiantil, prácticamente muda.

3. 6. 7. La Doctrina Especializada

A dieciséis meses de la publicación de ambas leyes LGPDPPSO y LFPDPPP, no se encontraron artículos en revistas indexadas mexicanas dedicados específicamente a la reforma. Lo que circula son comentarios de despachos (Greenberg Traurig, Basham, Garrigues, Hogan Lovells), publicaciones profesionales (IAPP) y análisis de sociedad civil. Por lo tanto, lo que se presenta a continuación son comentarios académicos que recaen en temas que siguen siendo pertinentes.

Diversos académicos mexicanos han planteado preocupaciones significativas sobre la eficacia del modelo tradicional de consentimiento, basado principalmente en la notificación, en el contexto de la IA y del procesamiento de datos a gran escala (Mendoza Enríquez, 2020; Quijano Decanini, 2022).

Como observa Quijano Decanini (2022) —analizando la legislación anterior, si bien buena parte de su crítica resulta trasladable a la vigente—, aunque las normas de protección de datos se han centrado históricamente en evitar la recolección secreta y en asegurar que el uso se alinee con el consentimiento informado, este paradigma flaquea ante los grandes volúmenes de microdatos de información, características de los entornos de datos contemporáneos. En tales escenarios, obtener un consentimiento significativo se convierte en una expectativa no razonable, dada la complejidad y el número de decisiones que se exigen al usuario.

A partir de las limitaciones que advierten en el modelo de notificación y consentimiento, estos autores plantean la necesidad de un cambio regulatorio orientado a la gobernanza del uso efectivo de los datos y al fortalecimiento de la rendición de cuentas corporativa. Quijano Decanini (2022), retomando a Lindu, propone que los esfuerzos se concentren en controlar

el uso de los datos en lugar de centrarse en los mecanismos de los avisos de privacidad y en la obtención del consentimiento. Ello implica discernir cuándo el consentimiento resulta primordial frente a aquellos supuestos en que la regulación del uso ofrece mejor protección, y exige de los proveedores que diseñen herramientas para que los usuarios puedan proteger y controlar su información personal.

Mendoza Enríquez (2020) complementa esta visión al sostener que, incluso cuando el consentimiento se obtiene conforme a la legislación aplicable, con frecuencia no llega a ser genuinamente informado, pues los usuarios pueden carecer de la capacidad o de los mecanismos necesarios para comprender plenamente la “dimensión de la autorización” otorgada. La crítica reformula el problema: no se trata de una cuestión de conocimiento, sino de comprensión, lo que exige un reequilibrio jurídico y ético que asegure la validez del consentimiento más allá del simple asentimiento. De ahí el llamado a que el principio de consentimiento evolucione, de modo que los individuos manifiesten realmente su voluntad respecto del uso de su información.

En este sentido, Tenorio Cueto sostiene que la autodeterminación informativa consiste en un ejercicio de libertad para entregar los propios datos, pero que ese ejercicio solo se configura si la persona conoce qué ocurrirá con ellos, comprende los alcances de la entrega y otorga su pleno consentimiento. Y cierra con una afirmación categórica: de no cumplirse esas condiciones, no cabe hablar propiamente de autodeterminación informativa (2019, p. 16).

Una de las observaciones críticas de mayor calado se refiere a la brecha entre el cumplimiento legal y la dimensión ética del tratamiento de datos personales, especialmente en contextos que involucran sistemas de IA. Mendoza Enríquez (2020) sostiene que el énfasis del modelo mexicano en el principio de información, expresado a través del aviso de privacidad, y en el consentimiento obtenido por esa vía resulta insuficiente para garantizar un tratamiento inherentemente ético. El consentimiento, aunque formalmente obtenido, difícilmente puede considerarse pleno cuando faltan mecanismos efectivos que permitan a los usuarios comprender de forma integral sus implicaciones.

Se configura así un escenario en el que las actividades de tratamiento pueden resultar lícitas conforme a derecho, pero no éticas (Mendoza Enríquez, 2020). El autor propone, en consecuencia, que los principios de lealtad y responsabilidad reciban mayor énfasis en relación con el consentimiento dentro del marco de protección de datos.

Lo expuesto sugiere que el enfoque actual del consentimiento informado, articulado a través de términos y condiciones, se encuentra superado (Mendoza Enríquez, 2020) y que debería otorgarse mayor peso a la responsabilidad de las partes de comportarse éticamente dentro de los parámetros de confianza propios de su relación, apoyada en una regulación flexible orientada a prevenir daños específicos. Esta referencia a la confianza es un tema clave y transversal en esta investigación.

La académica Sánchez Díaz (2024) señala la vulnerabilidad de la protección de datos personales ante la falta de un marco jurídico especializado frente al desarrollo tecnológico, con particular atención a la GenAI. La autora advierte la ausencia de sincronía entre la producción normativa y el cambio tecnológico, y sostiene que la opción gubernamental por la autorregulación de las empresas tecnológicas compromete la tutela de los derechos humanos y la seguridad de los datos personales, de donde concluye la necesidad imperativa de actualizar y robustecer los marcos normativos.

En una obra recién publicada, Tenorio Cueto señala que las tendencias regulatorias convergen hacia modelos basados en riesgo, señalando que “este enfoque corrige las limitaciones de los marcos centrados exclusivamente en el consentimiento, dando paso a mecanismos estructurales de rendición de cuentas” como el IA Act y el que propondremos en el siguiente capítulo, al referirnos a la gobernanza en términos de *accountability*.

3. 7. Análisis Comparativo de las Legislaciones de UE, EUA y México.

Los apartados precedentes han descrito por separado los ordenamientos europeo, estadounidense y mexicano. Corresponde ahora contrastarlos de manera sistemática, no para reiterar lo ya expuesto, sino para identificar los ejes en que convergen y divergen y extraer de ese contraste las consecuencias que interesan al objeto de esta investigación: determinar si el consentimiento informado del estudiantado puede sostener, por sí solo, la protección de sus datos personales frente a los sistemas de IA desplegados por las instituciones de educación superior.

La comparación se organiza en seis ejes —la arquitectura normativa, la posición del consentimiento en la estructura de licitud, el estándar de validez exigido, el tratamiento específico de la IA, la regulación de la cadena de proveedores y la arquitectura de garantía— y culmina en una síntesis tabular. Conviene advertir desde ahora que la divergencia decisiva entre los tres ordenamientos no se encuentra en el concepto de consentimiento, que resulta notablemente homogéneo, sino en la carga que cada uno de ellos hace recaer sobre esa figura y en los mecanismos que arbitra para sustituirla cuando resulta insuficiente.

3.7.1. La Arquitectura Normativa: Integración, Fragmentación y Dualidad

El ordenamiento europeo responde a un modelo integrado y transversal. Una norma general de aplicación directa —el RGPD— establece el estatuto común de todo tratamiento de datos personales, y una norma tecnológica sectorial —el AI Act— disciplina los sistemas de IA con independencia del sector en que se empleen, si bien identifica expresamente la educación como ámbito de alto riesgo. A ello se añade una capa de derecho blando interpretativo, singularmente las Directrices 5/2020 del Comité Europeo de Protección de Datos, que precisa el contenido de las obligaciones sin necesidad de reforma legislativa. El criterio ordenador

del ámbito de aplicación es el objeto: el dato personal y el sistema de IA, con independencia de quién los trate.

El modelo estadounidense se organiza, por el contrario, sobre una doble fragmentación. En el plano federal, la regulación es sectorial: FERPA para los registros educativos, HIPAA para la información sanitaria, GLBA para la financiera y COPPA para los datos de menores en servicios en línea. En el plano territorial, la ausencia de una ley general ha propiciado el patchwork de leyes estatales al que se hizo referencia en el apartado 3.5. De esta arquitectura se sigue una paradoja de considerable relevancia para las IES: la norma sustantivamente más protectora —la CPRA— define su ámbito subjetivo en torno a la noción de business de carácter lucrativo sujeto a umbrales de facturación o de volumen de datos, lo que deja fuera, con carácter general, a las instituciones sin fines de lucro, aunque alcanza plenamente a los proveedores tecnológicos que les prestan servicio; mientras que la norma que sí las alcanza —FERPA— es la más antigua y fue concebida para registros documentales en soporte físico y, como se verá, es también la menos exigible.

El ordenamiento mexicano introduce un tercer criterio, distinto de los dos anteriores: la dualidad por naturaleza jurídica del responsable. No es el objeto tratado ni el sector de actividad lo que determina el régimen aplicable, sino el carácter público o privado de la institución. Dos universidades que impartan los mismos programas empleen el mismo sistema de IA y traten idénticas categorías de datos quedan sujetas a regímenes diferentes, con distintas reglas sobre licitud, contratación de proveedores, transferencias y supervisión. Se trata, de los tres examinados, del criterio ordenador menos vinculado al riesgo efectivo del tratamiento.

3.7.2. La Posición del Consentimiento en la Estructura de Licitud

El segundo eje de comparación resulta más revelador que el primero. En el ordenamiento europeo el consentimiento es una de las seis bases de licitud del artículo 6.1 del RGPD y, en las relaciones caracterizadas por un desequilibrio de poder, ni siquiera la más adecuada: el artículo 7.4 y el considerando 43 advierten que el consentimiento no se considerará libremente prestado cuando exista una asimetría manifiesta entre el responsable y la persona titular. La consecuencia es que el ordenamiento europeo ha desplazado deliberadamente el centro de gravedad de la protección desde la voluntad individual hacia las obligaciones de diseño, las prohibiciones absolutas y las garantías de gobernanza.

El modelo estadounidense responde a una lógica distinta, pero conduce a un resultado convergente. FERPA formula el consentimiento previo y por escrito como regla para la divulgación de información personal identificable, pero lo acompaña de un catálogo tan amplio de excepciones —información de directorio, funcionarios escolares con interés educativo legítimo, estudios realizados por cuenta de la institución— que, en la práctica, la mayor parte del tratamiento ordinario ocurre fuera del ámbito del consentimiento y se

legítima por habilitación institucional. La CPRA, por su parte, invierte el planteamiento: el tratamiento se presume admisible y la protección se ejerce mediante mecanismos de exclusión voluntaria, reservando el consentimiento previo a los supuestos de menores de edad. Ni una ni otra sostienen realmente la protección sobre el acto de voluntad del titular.

El ordenamiento mexicano es, de los tres, el que conserva con mayor firmeza el discurso del consentimiento como regla de licitud —al menos en el régimen de los particulares— y, simultáneamente, el que menos instrumentos alternativos ofrece. La admisión del consentimiento tácito como modalidad ordinaria produce además un efecto que conviene nombrar con precisión: bajo la apariencia de un modelo de autorización previa opera, en realidad, un mecanismo de oposición posterior, funcionalmente próximo al *opt-out* estadounidense, pero sin las garantías procedimentales que aquél incorpora. De ello resulta que el ordenamiento mexicano mantiene la retórica del consentimiento habiendo renunciado a sus condiciones de eficacia.

3.7.3. El Estándar de Validez del Consentimiento

El contraste entre los tres estándares de validez arroja un resultado inequívoco. El RGPD exige que el consentimiento sea libre, específico, informado e inequívoco, manifestado mediante declaración o acción afirmativa clara; excluye el silencio, la inacción y las casillas premarcadas; impone la granularidad por finalidades; atribuye al responsable la carga de demostrar que lo obtuvo; y exige que su retirada sea tan sencilla como su otorgamiento. Las Directrices 5/2020 traducen estas exigencias en obligaciones auditables. La CPRA incorpora una definición prácticamente equivalente y añade tres exclusiones expresas de notable valor operativo: la aceptación de unas condiciones generales que describan conjuntamente el tratamiento y otra información no relacionada no constituye consentimiento; tampoco lo constituyen actos como pasar el cursor sobre un contenido, silenciarlo, pausarlo o cerrarlo; y el acuerdo obtenido mediante patrones oscuros carece igualmente de validez.

La definición mexicana omite el requisito de que el consentimiento sea inequívoco, admite la modalidad tácita como regla general y no incorpora ninguna de las precisiones anteriores: no prohíbe las casillas premarcadas, no define el patrón oscuro, no exige granularidad por finalidad ni establece una regla de equivalencia entre la facilidad de otorgamiento y la de retirada. En un entorno de interfaces digitales —que es, precisamente, el entorno en que operan los sistemas de IA— tales omisiones no son de detalle: son exactamente las que el legislador californiano estimó necesario colmar.

La comparación no debe conducir, sin embargo, a una conclusión unívoca. Existe al menos un extremo en que el ordenamiento mexicano resulta más exigente que el europeo: la reutilización de los datos para una finalidad distinta de la prevista en el aviso de privacidad requiere nuevo consentimiento sin que quepa invocar test de compatibilidad alguno, mientras que el artículo 6.4 del RGPD admite el tratamiento ulterior cuando la nueva finalidad resulte

compatible con la original. En el sector público mexicano la exigencia se acumula, pues el artículo 12 de la LGPDPPSO requiere además la atribución normativa que legitime el nuevo tratamiento. La reforma mexicana de 2025 fortaleció, por tanto, el principio de finalidad, aun cuando dejara intacta la debilidad del estándar de expresión de la voluntad.

3.7.4. El tratamiento normativo de la IA

Es en este eje donde la divergencia alcanza su máxima amplitud. La Unión Europea es el único de los tres ordenamientos que cuenta con una norma específica sobre IA, construida sobre una clasificación por niveles de riesgo que sitúa expresamente en la categoría de alto riesgo los sistemas empleados para determinar el acceso o la admisión, evaluar los resultados del aprendizaje, valorar el nivel educativo adecuado y supervisar el comportamiento del estudiantado durante las evaluaciones. A ello se añaden una prohibición absoluta —la inferencia de emociones en centros educativos, indisponible por consentimiento—, un deber de transparencia aplicable desde el 2 de agosto de 2026 y una obligación de alfabetización en IA exigible desde el 2 de febrero de 2025. El aplazamiento del régimen de alto riesgo operado por el Reglamento (UE) 2026/1744 no afecta a estos tres últimos extremos.

Estados Unidos carece de norma federal en la materia. La regulación más próxima procede de las reglas sobre tecnología de decisión automatizada y evaluaciones de riesgo adoptadas por la *California Privacy Protection Agency*, aplicables desde el 1 de enero de 2026, cuyo ámbito subjetivo replica, no obstante, las limitaciones ya señaladas. FERPA, concebida cinco décadas antes, no contempla la cuestión en modo alguno.

El ordenamiento mexicano no dispone de norma alguna sobre IA, y ni la LFPDPPP ni la LGPDPPSO emplean la expresión en su articulado. El único asidero normativo es el derecho de oposición frente al tratamiento automatizado del artículo 26, fracción II, de la LFPDPPP, cuya comparación con el artículo 22 del RGPD resulta ilustrativa de la distancia entre ambos modelos. El precepto europeo formula una prohibición general de las decisiones basadas únicamente en tratamiento automatizado, con excepciones tasadas y garantías correlativas: derecho a obtener intervención humana, a expresar el punto de vista y a impugnar la decisión. El precepto mexicano configura, en cambio, un derecho de oposición que la persona titular debe ejercer una vez iniciado el tratamiento, sin derecho correlativo a la intervención humana, a conocer la lógica aplicada ni a impugnar el resultado, y sometido a una excepción de contornos amplios cuando el tratamiento resulte necesario para el cumplimiento de una obligación legal del responsable.

La diferencia no es de grado sino de naturaleza: donde el ordenamiento europeo dispone una garantía estructural que opera ex ante, el mexicano ofrece una facultad individual, reactiva y de ejercicio incierto.

3.7.5. La Cadena de Responsabilidad

Los tres ordenamientos coinciden en un punto que resulta central para la tesis que aquí se sostiene: el momento jurídicamente decisivo para la protección de los datos del estudiantado no es aquel en que la institución recaba el consentimiento, sino el anterior, en que contrata al proveedor del sistema. El artículo 28 del RGPD impone un contrato de encargo con contenido mínimo tasado y sujeta la subcontratación a autorización; el AI Act distribuye obligaciones a lo largo de la cadena de valor y atribuye al responsable del despliegue —posición que ocupa la universidad que implanta el sistema— deberes propios de supervisión humana, vigilancia e información. FERPA condiciona la excepción del *school official* a que el proveedor actúe bajo el control directo de la institución y limite el uso al interés educativo legítimo. La CPRA distingue entre *service provider*, *contractor* y *third party* e impone cláusulas contractuales que impiden al proveedor destinar los datos a finalidades propias, con la consecuencia de que su incumplimiento reconfigura la comunicación como venta o cesión.

En México, como se expuso en el apartado 3.6.4, la LGPDPPSO desarrolla un régimen detallado de la relación entre responsable y encargado, incluida la contratación por adhesión de servicios de cómputo en la nube y la requalificación jurídica del encargado que decide por sí mismo sobre el tratamiento. La LFPDPPP carece por completo de ese desarrollo y, a la fecha, del reglamento que debía expedirse dentro de los noventa días siguientes a su entrada en vigor. La asimetría produce un efecto contraintuitivo: la institución privada, aparentemente sujeta a un régimen más flexible, es la que queda desprovista de herramientas jurídicas de rango legal para exigir garantías a sus proveedores.

3.7.6. La Arquitectura de Garantía

El último eje de comparación atiende a la instancia encargada de hacer exigibles los derechos reconocidos, y confirma que la eficacia de una garantía depende menos de su formulación que de su arquitectura institucional. El modelo europeo descansa sobre autoridades nacionales independientes coordinadas por el Comité Europeo de Protección de Datos, con potestad sancionadora de alcance disuasorio y sometidas a control judicial. La independencia de estas autoridades no es un rasgo meramente formal: fue precisamente el dictamen conjunto del Comité Europeo de Protección de Datos y del Supervisor Europeo el que advirtió sobre los riesgos del paquete Digital Omnibus (EDPB & EDPS, 2026), en un ejercicio de crítica institucional que sólo un órgano externo e independiente puede desempeñar.

El modelo estadounidense ofrece, en este punto, el contraste más agudo entre exigencia formal y exigibilidad real. FERPA se aplica exclusivamente por vía administrativa, a través del Student Privacy Policy Office, y su sanción última —el retiro de los fondos federales— nunca ha sido impuesta. El régimen californiano corrige parcialmente esta debilidad mediante una agencia especializada dotada de potestad sancionadora, pero sólo respecto de los sujetos comprendidos en su ámbito de aplicación.

En México, la reforma de 2024-2025 suprimió la autoridad garante independiente y redistribuyó sus competencias: la supervisión de los particulares corresponde ahora a una dependencia del Ejecutivo Federal y la de los sujetos obligados al órgano interno de control de cada institución. De ello resulta la inversión ya señalada en el apartado 3.6.5: el régimen sustantivamente más exigente —el del sector público— queda confiado a una supervisión interna, mientras que el más permisiva conserva, al menos formalmente, una autoridad externa. La intensidad de la exigencia sustantiva y la de la garantía institucional operan en sentido contrario.

La tabla siguiente sintetiza los criterios de comparación empleados en este apartado, con el propósito de facilitar la lectura conjunta de los tres regímenes y de servir de base al análisis conclusivo que cierra el capítulo.

Tabla 2. Análisis comparado de los regímenes de consentimiento informado y protección de datos aplicables a las instituciones de educación superior: Unión Europea, Estados Unidos y México.

Criterio de comparación	Unión Europea	Estados Unidos	México
Modelo regulatorio	Integrado y transversal: un reglamento general de aplicación directa (RGPD), complementado por una norma tecnológica sectorial (AI Act).	Fragmentado en doble eje: sectorial en el plano federal (FERPA, HIPAA, GLBA, COPPA) y territorial en el estatal (patchwork de privacidad).	Dual en función de la naturaleza jurídica del responsable: LFPDPPP para particulares y LGPDPPSO para sujetos obligados.
Criterio ordenador del ámbito de aplicación	El objeto: el dato personal y el sistema de IA, con independencia de quién los trate.	El sector de actividad y la residencia de la persona titular; en la CPRA, además, el carácter lucrativo y los umbrales del responsable.	La naturaleza pública o privada del responsable, con independencia del sector y del riesgo del tratamiento.
Norma general de datos aplicable a la IES	RGPD (Reglamento UE 2016/679), aplicable a toda institución.	Ninguna de alcance federal. La CPRA no alcanza, con carácter general, a las IES sin fines de lucro, aunque sí a sus proveedores tecnológicos.	LFPDPPP (2025) o LGPDPPSO (2025), según la titularidad de la institución.
Norma sectorial educativa	No existe norma específica; el sector se aborda desde el AI Act (anexo III, punto 3, y considerando 56).	FERPA (1974), aplicable a toda institución que reciba fondos federales del Departamento de Educación.	Ley General de Educación: prácticamente muda en materia de consentimiento estudiantil e inaplicable al alumnado mayor de edad.
Norma específica sobre IA	Reglamento (UE) 2024/1689, modificado por el Reglamento (UE) 2026/1744, con enfoque basado en el riesgo.	Sin norma federal. Reglas de la CPPA sobre tecnología de decisión automatizada (ADMT) y evaluaciones de riesgo, aplicables desde el 1 de enero de 2026.	Ninguna. Las leyes vigentes no emplean en su articulado la expresión “IA”.
Definición legal de consentimiento	Manifestación libre, específica, informada e inequívoca, mediante declaración o acción afirmativa clara (art. 4.11).	En la CPRA, definición equivalente a la europea. FERPA exige consentimiento previo, escrito, fechado y firmado para la divulgación.	Manifestación libre, específica e informada (art. 2, fracc. IV LFPDPPP; art. 3, fracc. VIII LGPDPPSO). No se exige que sea inequívoca.
Modalidad ordinaria de expresión	Expreso, mediante acción afirmativa. Se excluyen el silencio, la inacción y las casillas premarcadas (considerando 32).	FERPA: escrito. CPRA: exclusión voluntaria (<i>opt-out</i>) como regla, con consentimiento previo solo para menores de dieciséis años.	Tácito como regla general: basta poner el aviso a disposición y que la persona titular no manifieste oposición (art. 7 LFPDPPP; art. 15 LGPDPPSO).
Estándar reforzado para datos sensibles	Prohibición general de tratamiento salvo excepciones tasadas; el consentimiento ha de ser explícito (art. 9).	CPRA: derecho a limitar el uso y la divulgación de la información personal sensible.	Consentimiento expreso y por escrito, mediante firma autógrafa, electrónica o mecanismo de autenticación equivalente.

Reutilización para finalidad distinta	Admisible si la nueva finalidad es compatible, conforme al test del art. 6.4; en otro caso, se requiere nueva base de licitud.	CPRA: limitación a finalidades razonablemente compatibles con el contexto de la recogida y divulgadas previamente.	Exige nuevo consentimiento, sin test de compatibilidad (art. 11 LFPDPPP). En el sector público, además, atribución normativa expresa (art. 12 LGPDPPSO).
Tratamiento del desequilibrio de poder	Reconocido expresamente: se presume la falta de libertad ante desequilibrio (art. 7.4 y considerando 43; Directrices 5/2020). Fundamenta el art. 5.1.f del AI Act.	No se formula como categoría general; opera de modo indirecto a través de las excepciones institucionales de FERPA.	No contemplado. La relación de matrícula no recibe tratamiento diferenciado alguno.
Decisiones automatizadas	Derecho a no ser objeto de decisiones basadas únicamente en tratamiento automatizado, con intervención humana, expresión del punto de vista e impugnación (art. 22).	Reglas de la CPPA sobre ADMT: derecho de exclusión, información previa sobre la lógica y evaluación de riesgo obligatoria.	Derecho de oposición ejercitable ex post, sin derecho a intervención humana, a explicación ni a impugnación, y exceptuado ante obligación legal (art. 26, fracc. II).
Evaluación de impacto o de riesgo previa	Evaluación de impacto obligatoria en tratamientos de alto riesgo (art. 35 RGPD) y sistema de gestión de riesgos para sistemas de alto riesgo (AI Act).	Evaluaciones de riesgo obligatorias conforme a las reglas de la CPPA, dentro de su ámbito subjetivo de aplicación.	No prevista con carácter obligatorio en ninguno de los dos regímenes.
Prohibiciones absolutas en el ámbito educativo	Prohibido inferir emociones en centros educativos, salvo fines médicos o de seguridad (art. 5.1.f). La prohibición es indisponible por consentimiento.	No existen prohibiciones absolutas equivalentes.	No existen prohibiciones absolutas.
Régimen del encargado y de la contratación de proveedores	Contrato obligatorio con contenido mínimo (art. 28 RGPD), autorización para subencargados y obligaciones propias del responsable del despliegue (art. 26 AI Act).	FERPA: excepción del school official, bajo control directo de la institución. CPRA: cláusulas obligatorias que impiden al proveedor destinar los datos a fines propios.	LGPDPPSO: régimen detallado (arts. 52 a 58), incluida la contratación de servicios en la nube. LFPDPPP: sin desarrollo legal y sin reglamento vigente.
Deber de informar los destinatarios de las transferencias	Obligatorio: destinatarios o categorías de destinatarios y transferencias internacionales previstas (arts. 13 y 14).	CPRA: obligación de informar las categorías de terceros a quienes se divulga la información personal.	LGPDPPSO: obligatorio (art. 21, fracciones VII y VIII). LFPDPPP: suprimido; solo se exige una cláusula de aceptación o rechazo (art. 35).
Alfabetización en IA	Obligación de garantizar un nivel suficiente de alfabetización en IA (art. 4 AI Act, exigible desde el 2 de febrero de 2025) y considerando 56.	No prevista.	No prevista.
Autoridad de control	Autoridades nacionales independientes y Comité Europeo de Protección de Datos; autoridades de	FERPA: Student Privacy Policy Office del Departamento de Educación. CPRA: California Privacy	INAI extinguido. Particulares: Secretaría Anticorrupción y Buen Gobierno. Sujetos obligados: órgano interno

	vigilancia del mercado en materia de IA.	Protection Agency y Fiscalía General del Estado.	de control de la propia institución.
Exigibilidad individual	Reclamación ante la autoridad de control, tutela judicial efectiva y derecho a indemnización (arts. 77 a 82).	FERPA no confiere acción privada (Gonzaga University v. Doe, 536 U.S. 273, 2002); la sanción última es el retiro de fondos federales, nunca aplicada.	Procedimiento de protección de derechos ante la autoridad competente; en el sector público, ante el órgano interno de control del propio sujeto obligado.

Nota. Elaboración propia con apoyo de Gemini (Google, 2026), a partir del prompt "Haz una tabla que compare de los regímenes de consentimiento informado y protección de datos aplicables a las instituciones de educación superior: Unión Europea, Estados Unidos y México, de conformidad con las siguientes categorías: modelo regulatorio, criterio ordenador del ámbito de aplicación, norma general de datos aplicable a la IES, norma sectorial educativa, norma específica sobre IA, definición legal de consentimiento, modalidad ordinaria de expresión, estándar reforzado para datos sensibles, reutilización para finalidad distinta, tratamiento del desequilibrio de poder, decisiones automatizadas, evaluación de impacto o de riesgo previa, prohibiciones absolutas en el ámbito educativo, régimen del encargado y de la contratación de proveedores, deber de informar los destinatarios de las transferencias, alfabetización en IA, autoridad de control y exigibilidad individual," y del contenido normativo desarrollado por la autora.

De la lectura conjunta de la tabla se desprenden tres hallazgos que conviene enunciar con precisión. El primero es de convergencia conceptual: los tres ordenamientos definen el consentimiento en términos casi idénticos, de modo que la divergencia no se sitúa en el concepto, sino en la carga que cada uno hace recaer sobre él. El segundo es de divergencia funcional: la Unión Europea ha desplazado expresamente la protección desde la voluntad individual hacia las obligaciones de diseño, las prohibiciones absolutas y la gobernanza del despliegue; Estados Unidos opera ese desplazamiento de manera parcial, por vía contractual y de mercado; y México conserva un modelo centrado en el consentimiento sin haber articulado los instrumentos que habrían de sustituirlo. El tercero es de inversión entre exigencia y garantía: allí donde la exigencia sustantiva mexicana alcanza su punto más alto, la garantía institucional alcanza el más bajo.

La tecnología avanza más rápido de lo que legislan los parlamentos, y esa distancia no es un accidente pasajero: la propia Unión Europea, que sí llegó a aprobar una ley de IA, tuvo que aplazar su parte más exigente antes siquiera de que empezara a aplicarse. En México ni siquiera existe una iniciativa con horizonte previsible, y el reglamento que debía desarrollar la ley federal sigue sin publicarse. Esperar a la norma significa dejar sin gobierno justamente los años en que los sistemas se están instalando en las aulas.

A esa urgencia se suma una constatación de fondo: el consentimiento no puede cargar con todo el peso que se le ha atribuido. Autores como Solove (2013), Zeide (2016), Tenorio Cueto (2019), Mendoza Enríquez (2020) y Quijano Dacanini (2022) han mostrado desde tradiciones distintas que pedir a cada persona que lea, entienda y evalúe las consecuencias de cada autorización es pedirle algo que las condiciones reales no permiten. La conclusión común es que la protección debe desplazarse del acto de voluntad individual hacia las obligaciones de

quien usa los datos; y si la ley mexicana no efectúa ese desplazamiento, la única instancia que puede hacerlo es la IES misma.

Por lo anterior, en el capítulo siguiente se integran referentes internacionales de gobernanza y estándares de cumplimiento —como las normas ISO— para sustentar la propuesta del Modelo de Gobernanza para la Confianza, concebido como un marco integral para operacionalizar un consentimiento informado, afirmativo, específico, granular, verificable y revocable en la IES.

La combinación de regulación pública, autorregulación vinculante y gobernanza institucional puede contribuir a que la protección de los datos personales, mediante el consentimiento informado, deje de ser una obligación meramente formal y se convierta en una práctica organizacional continua, verificable y centrada en los derechos de la comunidad estudiantil.

Capítulo 4. Propuesta del Modelo de Gobernanza para la Confianza (MGC) para la incorporación adecuada de la IA en educación superior.

Como culminación de los objetivos específicos de esta investigación, y una vez analizados los impactos, principios éticos y marcos normativos en los capítulos precedentes, este apartado cumple con la cuarta línea de acción: proponer un modelo de gobernanza aplicable a las IES. El Modelo de Gobernanza para la Confianza (MGC) integra dimensiones de gobernanza, operación, pedagogía y formación, incorporando mecanismos de cumplimiento (*compliance*) y estándares internacionales para operacionalizar las exigencias éticas y jurídicas, fortaleciendo así el consentimiento informado en las instituciones mexicanas. Los capítulos anteriores permiten identificar una brecha en los enfoques actuales para abordar de manera integrada estos desafíos. A partir de esta brecha, se presenta la propuesta del MGC y su operacionalización mediante indicadores clave de desempeño (KPI), con el fin de fortalecer las prácticas institucionales y garantizar una protección efectiva de la privacidad en la era de la IA.

La Pontificia Academia para la Vida (2020) afirma que la ley, por sí sola, es insuficiente para garantizar la justicia algorítmica, de modo que la ética debe preceder y orientar su aplicación. En este sentido, la UNESCO (2021) sostiene que, en ausencia de marcos legales integrales, los principios éticos deben guiar el desarrollo y la gobernanza de los sistemas de IA. Así pues, mientras los marcos regulatorios continúan en evolución, la ética se convierte en un mecanismo orientador del uso responsable de la IA.

Sin embargo, la existencia de principios no garantiza su eficacia. Como señala Chan (2023) “hasta que los principios sean validados por la implementación en tiempo real de las tecnologías de IA, permanecerán principalmente de naturaleza predictiva y prescriptiva” (p. 5). Consideramos que los lineamientos genéricos centrados exclusivamente en *ética* si no se traducen en procedimientos, responsabilidades y capacidades operativas dentro de las universidades los hacen insuficientes (Chan, 2023). En consecuencia, es imperativo descender hacia niveles de operacionalización concreta. Así lo señala la OCDE (2022) quien propone mecanismos éticos de gobernanza.

Para transitar de los principios abstractos y las directrices hacia los mecanismos éticos de gobernanza es necesario adoptar políticas y modelos de implementación. En el capítulo segundo se revisaron algunos marcos como las Directrices éticas para una IA fiable (UE) y el *AI Risk Management Framework* del NIST, EE. UU., los cuales permiten estructurar la gobernanza de manera iterativa y documentable (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023). En este capítulo presentaremos el “Marco de políticas de educación ecológica de IA” (Chan, 2023) y mencionaremos la ISO, entendidas no como fuentes de obligaciones jurídicas, sino como instrumentos de autorregulación y el cumplimiento normativo (*compliance*), que contribuyen a operacionalizar el cumplimiento

de los principios legales mediante sistemas de gestión, controles documentados y mecanismos de mejora continua.

En conjunto, estos marcos aportan elementos esenciales para una gobernanza responsable de la IA; sin embargo, ninguno integra de manera articulada las dimensiones jurídicas, ética, tecnológica y pedagógica necesarias para operacionalizar el consentimiento informado en las IES. Esta limitación justifica la formulación del Modelo de Gobernanza para la Confianza que se presenta en este capítulo.

4. 1. El Marco de Políticas de Educación Ecológica de IA⁹

Su propuesta deriva de la teoría ecológica, aplicada a la política educativa. El modelo de Chan se basa en que la integración de la IA en la educación es un fenómeno que involucra partes interconectadas: la Alta Dirección, los operadores, los docentes y los estudiantes.

La analogía con los sistemas biológicos enfatiza que la alteración de un componente (por ejemplo, la introducción de un nuevo algoritmo de aprendizaje personalizado) afecta a todo el sistema: la privacidad del estudiante, la carga laboral del docente, la equidad educativa. La propuesta ecológica convierte la amenaza de la IA en algo que se puede manejar dentro del sistema educativo.

Chan estructura su modelo ecológico en tres dimensiones concéntricas que operan en simbiosis: la dimensión de gobernanza, la dimensión operativa y tecnológica y la dimensión pedagógica, que se abordarán en los siguientes apartados.

Si bien el modelo ecológico de Chan constituye una de las propuestas más completas para orientar la incorporación responsable de la IA en educación superior, su propósito no consiste en desarrollar un modelo específico para la gobernanza del consentimiento informado ni en operacionalizar los mecanismos jurídicos y técnicos que este requiere. En consecuencia, su principal aporte para esta investigación radica en proporcionar la arquitectura institucional sobre la cual pueden integrarse elementos normativos, tecnológicos y pedagógicos adicionales.

⁹ La investigación de Chan (2023) parte de un análisis de la situación y una revisión crítica de las respuestas institucionales de alumnos de grado, posgrado y docentes de Hong Kong tras el lanzamiento de Chat GPT. Utilizando una metodología cuantitativa y cualitativa, concluye que existe un vacío de políticas educativas adecuadas para hacer frente a la incorporación de IA generativa (GenAI). A partir de este diagnóstico, Chan sistematizó las necesidades de los múltiples actores educativos: estudiantes, docentes, personal administrativo y empleadores.

4. 1. 1. La Dimensión de Gobernanza

Esta dimensión se centra en asegurar que los distintos miembros del ecosistema de la educación superior comprendan y aborden los desafíos éticos asociados con las tecnologías de IA, fomentando el uso responsable y ayudando a mantener la confianza dentro de la comunidad universitaria, buscando garantizar que estudiantes y docentes puedan interactuar de una manera ética con la IA (Chan, 2023).

La Alta Dirección debe garantizar que la IA se utilice de manera responsable y ética, fomentando un entorno de aprendizaje que sea justo, equitativo e inclusivo (Chan, 2023). Es la responsable de desarrollar y hacer cumplir políticas, directrices y procedimientos que aborden las preocupaciones éticas en torno al uso de la IA en la educación. Estos incluyen la integridad académica, la privacidad de los datos, la transparencia, la rendición de cuentas y la seguridad.

La gobernanza efectiva no puede imponerse únicamente de arriba hacia abajo (*top-down*); por el contrario, debe nutrirse de una comprensión holística de los actores, por lo que es clave considerar las otras dimensiones: operativa y tecnológica, así como la docente (Chan, 2023).

La dimensión de gobernanza protege el principio de responsabilidad (*accountability*), legalidad y gobernanza del tratamiento de datos cuando intervienen sistemas de IA en decisiones académicas.

Desde la perspectiva de esta investigación, esta dimensión resulta especialmente relevante porque constituye el espacio institucional desde el cual pueden articularse los principios de responsabilidad (*accountability*), legalidad y gobernanza del tratamiento de datos cuando intervienen sistemas de IA en decisiones académicas.

4. 1. 2. La Dimensión Operativa y Tecnológica

Esta dimensión se centra en la necesidad de un monitoreo, evaluación y apoyo continuos para asegurar la implementación efectiva y equitativa de las tecnologías de IA. Se enfatiza la importancia de la mejora continua y la adaptación, considerando que las IES deben ajustar sus estrategias de integración tecnológica en respuesta al entorno cambiante y evolutivo. La autora propone dos tareas distintas en esta dimensión: la gestión y el mantenimiento de las herramientas de IA en la educación, y la capacitación en el uso de estos sistemas o *alfabetización en IA (AI literacy)* (Chan, 2023). En esta dimensión se busca proteger el derecho de seguridad, confidencialidad e integridad de la información, junto con el principio de protección de datos desde el diseño y por defecto.

Para cumplir estas distintas tareas los actores en esta dimensión son dos distintas áreas: por un lado, el centro de formación docente y, por otro, el equipo de tecnologías o transformación

digital. El papel de la primera área es proporcionar capacitación y apoyo para el personal académico; en cambio, el segundo se responsabiliza de asegurar el funcionamiento adecuado de las herramientas de IA y abordar cualquier problema técnico que pueda surgir (Chan, 2023).

4. 1. 3. La Dimensión Pedagógica

En la dimensión pedagógica, Chan destaca la importancia de ajustar los métodos de enseñanza y las formas de evaluar, teniendo en cuenta las habilidades crecientes de la IA. Esto ayuda a preparar a los estudiantes para un entorno laboral que se automatiza cada vez más (Chan, 2023). Proponemos que las tecnologías de IA se aprovechen para mejorar los resultados de aprendizaje y desarrollar el pensamiento crítico, la creatividad y otras habilidades esenciales, en lugar de socavar la integridad académica. En este contexto, Chan señala que se debe modificar la definición de integridad académica y evaluación en la era de la IA generativa, asegurando que el uso de herramientas tecnológicas potencie, en lugar de sustituir, el pensamiento crítico (Chan, 2023).

Los docentes son los iniciadores de la dimensión pedagógica, ya que son quienes diseñan e implementan los planes de estudio, actividades y evaluaciones que utilizan tecnologías de IA. Los docentes necesitarán tener la experiencia para determinar cómo la IA puede apoyar y mejorar las experiencias de aprendizaje de los estudiantes (Chan, 2023). Al mismo tiempo, deben poseer la conciencia y el conocimiento necesarios para educar a los estudiantes sobre los riesgos potenciales asociados con el uso de la IA generativa en el aprendizaje, particularmente en evaluaciones y tareas, tales como el plagio o la invención de datos. Los docentes, además de supervisar el uso de la IA, también deben fomentar el uso ético de la IA y desarrollar tareas de evaluación que requieran pensamiento crítico y analítico. Esto es necesario para asegurar que las herramientas de IA se utilicen de una manera pedagógicamente sólida y que mejore los resultados de aprendizaje de los estudiantes. En definitiva, los docentes deben desarrollar competencias holísticas en los estudiantes (Chan, 2023).

En la dimensión pedagógica se busca salvaguardar la calidad educativa, la libertad académica y principio de proporcionalidad cuando la IA implique monitoreo, analítica o automatización de decisiones sobre el estudiantado.

La conexión de las tres dimensiones —gobernanza, pedagógica y operacional— (Chan, 2023) ayuda a transitar de la estrategia hacia generar confianza en la IES al asegurar que haya consistencia entre los valores, las normas y las prácticas diarias (OCDE, 2022).

El aporte central del modelo de Chan para esta investigación es que permite situar el consentimiento informado dentro de un ecosistema institucional, y no como un acto aislado. Esta visión ecosistémica coincide con lo desarrollado en el capítulo 2 respecto a que la ética

de la IA requiere condiciones estructurales para generar confianza en la comunidad educativa. Ese enfoque sistémico refuerza la idea de que el consentimiento surge de la coherencia entre gobernanza, operación y pedagogía. Dicha coherencia es la base para construir la confianza institucional.

El modelo ecosistémico de Chan ofrece la base conceptual necesaria para entender el consentimiento informado como un proceso que es dinámico, relacional y transparente. Este enfoque permite incluir otros modelos y enfoques que se revisaron en el capítulo 3 de esta investigación.

4. 2. Integración de Marcos complementarios al Modelo Ecológico de Chan

La selección de estos marcos no responde a una acumulación de referencias, sino a su complementariedad funcional. Mientras el modelo ecológico de Chan proporciona la estructura institucional para la gobernanza de la IA en educación superior, las Directrices Éticas para una IA Fiable de la Unión Europea, el AI RMF del NIST y el modelo ACE incorporan componentes relacionados con la gestión del riesgo, la rendición de cuentas, la transparencia, la protección de derechos y el papel del docente. En conjunto, estos marcos fortalecen la capacidad del modelo ecológico para operacionalizar el consentimiento informado dentro de las instituciones de educación superior.

En primer lugar, observamos cómo el modelo de Chan (2023) se complementa con los marcos de la UE y NIST. Mientras Chan (2023) establece el ecosistema institucional (gobernanza, pedagogía y operaciones) necesario para alojar el consentimiento, los marcos UE/NIST proporcionan un proceso iterativo de gobernanza y gestión de riesgos —delimitar contextos, medir impactos, establecer controles y revisar decisiones— necesario para ejecutarlo con trazabilidad y rendición de cuentas. Al exigir transparencia, supervisión humana y control sobre datos y decisiones, el consentimiento se vuelve un proceso revisable, con información actualizada y vías efectivas de oposición y reparación (Comisión Europea, 2019; NIST, 2023).

De esta convergencia de los tres modelos (Comisión Europea, 2019; NIST, 2023; Chan, 2023), se desprende la premisa central de este capítulo: el consentimiento informado en la era de la IA es más que un documento legal firmado al inicio de un curso académico. Debe ser un proceso educativo y dialógico continuo, sostenido por una infraestructura de gobernanza ética y validado por la alfabetización digital del docente y el estudiante.

El modelo ACE (2023) complementa a Chan (2023). Ambas propuestas comparten una visión humanista de la enseñanza. La premisa central de ACE es que el docente debe estar en el centro del proceso educativo. Chan exige abordar valores esenciales como las relaciones profesor-estudiante y asigna a los docentes el liderazgo pedagógico.

El modelo de Chan actúa como el mecanismo estructural que hace posible el principio del modelo ACE de centrar al educador. Por su parte, este modelo ACE señala que en el proceso educativo hay tres momentos en los cuales el educador debe mantener el control: la enseñanza directa; la preparación y reflexión sobre la enseñanza; y el diseño, selección y evaluación de herramientas. Estos momentos son los que se dan en la dimensión pedagógica señalada por Chan.

Esta convergencia teórica sienta las bases para formular una propuesta propia, orientada no sólo al cumplimiento normativo, sino a la construcción activa de confianza institucional. La conexión entre estructura e iteración es clave. El consentimiento informado es un proceso institucional revisable, coherente y basado en principios éticos (Chan, 2023; Comisión Europea, 2019; NIST, 2023; U. S. Department of Education, 2023).

La integración de estos elementos constituye el fundamento conceptual del Modelo de Gobernanza para la Confianza. No obstante, para traducir estos principios en procesos institucionales verificables y sostenibles resulta necesario incorporar mecanismos de gestión y cumplimiento, aspecto que se desarrolla en el apartado siguiente.

4.3. Gobernanzas de las IES Mexicanas y Estándares ISO.

Los marcos analizados en el apartado anterior permiten identificar los principios que deben orientar la gobernanza del consentimiento informado en entornos de IA. Sin embargo, la existencia de principios no garantiza por sí misma su aplicación consistente dentro de las instituciones. Para que estos puedan implementarse, evaluarse y mantenerse en el tiempo, resulta necesario incorporar estándares de gestión que traduzcan dichos principios en procesos, responsabilidades y mecanismos verificables. En este contexto, las normas ISO constituyen el puente entre la gobernanza conceptual y la gobernanza operativa.

4.3.1. Gobernanza Institucional en las IES Mexicanas

En los últimos años, diversas IES han comenzado a desarrollar marcos, políticas y lineamientos institucionales para orientar la incorporación de la IA en la docencia, la investigación y la gestión universitaria. Estas iniciativas reflejan una transición desde recomendaciones de uso responsable hacia esquemas más amplios de gobernanza institucional.

El diagnóstico del Observatorio Interinstitucional de IA en la Educación Superior en México (OIIAES, 2026), levantado sobre 161 IES de 31 entidades federativas, confirma a escala nacional la tendencia que esta investigación identificó en su muestra intencional: el 57.57% de las instituciones encuestadas carece de cualquier documento normativo sobre el uso de la IA, mientras que aproximadamente una cuarta parte cuenta ya con entre uno y tres instrumentos oficiales. Esta convergencia entre una muestra cualitativa de siete instituciones y un diagnóstico de cobertura nacional refuerza la validez externa del hallazgo central del

apartado 4.3.1: la ausencia de normativa no es un rasgo aislado, sino la condición predominante del sistema mexicano de educación superior.

El mismo diagnóstico revela una brecha aún mayor en los mecanismos de verificación que en la existencia de políticas: el 86.96% de las IES encuestadas (140 de 161) no cuenta con ningún mecanismo para evaluar el uso de la IA dentro del aula, y el 64.06% (98 instituciones) no ha desarrollado estrategia alguna para conocer cómo su propio estudiantado utiliza estas herramientas (OIIAES, 2026). Esta ausencia de trazabilidad institucional —no sólo normativa, sino operativa— es precisamente el punto ciego que el MGC busca cerrar mediante indicadores verificables de consentimiento, confianza y auditabilidad.

Para presentar ejemplos concretos se han analizado siete IES mexicanas. La Tabla 3 presenta, de manera sintética, algunos de los principales referentes nacionales que permiten contextualizar la presente propuesta del Modelo de Gobernanza para la Confianza (MGC). Aunque dichos documentos comparten elementos con este modelo, ninguno integra de forma articulada las dimensiones jurídicas, de gobernanza, operativa, pedagógica y formativa orientadas específicamente al consentimiento informado y la protección de los datos personales.

Tabla 3 *Instrumentos de gobernanza de la IA en siete instituciones mexicanas de educación superior (2023–2026)*

Institución	Instrumento(s)	Año	Instancia emisora	Destinatario declarado	Naturaleza del instrumento	Jerarquía y alcance	Unidad de aplicación	Principios enunciados en fuentes consultadas
Tecnológico de Monterrey	Lineamientos para el uso ético de la IA (versión estudiantes y versiones docentes); Política de uso de herramientas de IA	2023–2024	Institucional / Integridad Académica	Estudiantes, docentes, colaboradores	Documentos segmentados por audiencia, con anexos operativos, y política de personal	Institucional; segmentado por destinatario	Actividad de aprendizaje declarada en syllabus	Integridad académica, atribución de autoría, propiedad intelectual, pensamiento crítico, dignidad humana
UAM	Decálogo de ética para el uso de la IA; Lineamientos éticos Unidad Azcapotzalco	2026	Comisión académica designada por Rectoría General; Consejo Académico de unidad	Comunidad universitaria	Recomendaciones preliminares + lineamientos por unidad	Subsidiario; delega en consejos divisionales	Unidad y división académica	Remite a Código de Ética institucional (2022), UNESCO (2021), Principios de Chapultepec
UNAM	Consejo Coordinador de IA (CCOIA) y Coordinación General	2026	Rectoría / Secretaría de Desarrollo Institucional	Institución (lineamientos en elaboración)	Órgano de gobernanza	Institucional transversal	Institucional	Ética, protección de datos, dignidad humana, propiedad intelectual, perspectiva humanista
BUAP	Consejo Académico de Gobernanza de IA; Agencia AURA	2026	Acuerdo de Rectoría	Institución	Órgano consultivo-estratégico + agencia operativa	Institucional	Institucional	Ética tecnológica, gobernanza digital, responsabilidad social
UACJ	Lineamientos para el uso ético de la IA en investigación	2025	Comité de Ética institucional	Personal investigador	Lineamientos procedimentales	Sectorial (función de investigación)	Producto de investigación / publicación	Integridad académica, transparencia, responsabilidad, declaración de uso y de prompts

Institución	Instrumento(s)	Año	Instancia emisora	Destinatario declarado	Naturaleza del instrumento	Jerarquía y alcance	Unidad de aplicación	Principios enunciados en fuentes consultadas
Universidad Panamericana	Marco general para el uso ético de sistemas de IA	2025	Rectoría General (elaborado por el comité institucional de uso ético de IA)	Docentes y estudiantes	Documento único de principios orientado a la formación de criterio, con desarrollo por fase del ciclo didáctico	Institucional; documento único aplicable a todos los campus	Curso y evidencia de aprendizaje	Justicia, transparencia, protección de datos, no discriminación, control humano, rendición de cuentas
UASLP	Sitio de lineamientos sugeridos; Grupo Universitario de IA (GUIA)	2023 / 2026	Secretaría Académica; acuerdo de Rectoría (Centro de Emprendimiento e Innovación)	Docentes y estudiantes	Recomendaciones no vinculantes + órgano colegiado	Orientativo; decisión en el docente	Curso	Soberanía tecnológica, transparencia, protección de la privacidad, impacto social

Nota. Elaboración propia con apoyo de Gemini (Google, 2026), a partir del prompt "Haz una tabla que compare los instrumentos de gobernanza de la IA en siete instituciones mexicanas de educación superior (2023–2026), conforme a las siguientes categorías: instrumento(s), año, instancia emisora, destinatario declarado, naturaleza del instrumento, jerarquía y alcance, unidad de aplicación, y principios enunciados en fuentes consultadas a partir del texto que te proporciono". Las columnas de principios y herramientas recogen únicamente formulaciones documentadas en las fuentes consultadas y no suponen exhaustividad respecto del contenido íntegro de cada instrumento; la indicación "no documentadas" señala ausencia en las fuentes consultadas, no inexistencia. UAM = Universidad Autónoma Metropolitana; UNAM = Universidad Nacional Autónoma de México; BUAP = Benemérita Universidad Autónoma de Puebla; UACJ = Universidad Autónoma de Ciudad Juárez; UASLP = Universidad Autónoma de San Luis Potosí.

En conjunto, los instrumentos analizados muestran una convergencia en principios fundamentales como la transparencia, la protección de los datos personales, la supervisión humana, la responsabilidad y la no discriminación. No obstante, presentan diferencias importantes en su alcance institucional y en los mecanismos propuestos para su implementación. Mientras algunos privilegian la integridad académica, otros enfatizan la gobernanza tecnológica, la formación ética o la investigación responsable. En general, se observa una limitada incorporación de mecanismos verificables para la evaluación de riesgos, el seguimiento institucional, la gestión de proveedores tecnológicos, la clasificación de sistemas de IA y la operacionalización del consentimiento informado. Estas áreas representan una oportunidad para avanzar hacia modelos de gobernanza más integrales en las instituciones de educación superior.

4.3.2. Estándares ISO

El reconocimiento normativo de derechos y deberes no garantiza, por sí mismo, su observancia práctica. Por eso, el Derecho actual ha cambiado hacia modelos de *compliance* o gestión del cumplimiento normativo. Estos modelos buscan convertir las obligaciones legales en procesos organizacionales que se pueden verificar, con el objetivo de prevenir riesgos, documentar acciones y mostrar que se cumplen las normas correspondientes (Parker & Nielsen, 2009; Voigt & von dem Bussche, 2017). De manera sintética, y en consonancia con la postura de la *World Compliance Association*, el *compliance* entraña la gestión proactiva de los riesgos legales y operativos en el seno de las organizaciones.

En las IES, caracterizadas por la gestión masiva de datos personales de estudiantes, docentes e investigadores, este enfoque encuentra su fundamento medular en el principio de responsabilidad demostrada (*accountability*). Siguiendo este principio, las universidades no solo deben cumplir con las obligaciones de las normativas legales de protección de datos, sino también demostrar, con pruebas claras, que este cumplimiento es una parte esencial de su modelo de gobernanza. Como subraya Lachaud (2017, 2020), este imperativo adquiere una dimensión paradigmática bajo el marco del Reglamento General de Protección de Datos (RGPD) y regulaciones análogas.

Desde esta perspectiva, los estándares internacionales desarrollados conjuntamente por la *International Organization for Standardization* y la *International Electrotechnical Commission* (ISO/IEC*¹⁰) constituyen instrumentos técnicos de *compliance* que complementan el ordenamiento jurídico universitario. Las normas ISO no reemplazan las obligaciones de la ley, sino que las hacen más efectivas a través de sistemas de gestión, controles documentados y procedimientos que se pueden auditar. Así, transforman principios normativos —como la legalidad, la transparencia, la seguridad y el consentimiento informado— en prácticas organizativas que se pueden verificar de manera objetiva en el entorno educativo.

Uno de los principales referentes en esta materia es la norma ISO 27701 del 2019, la cual extiende los sistemas de gestión de seguridad de la información (ISO 27001 e ISO 27002) mediante la incorporación de requisitos específicos para un Sistema de Gestión de Información de Privacidad (*Privacy Information Management System*, PIMS). En el ámbito académico, el objetivo es establecer un sistema de gestión que ayude a identificar los riesgos relacionados con el manejo de datos de la comunidad universitaria, asignar responsabilidades, manejar incidentes y asegurar el cumplimiento de las reglas (ISO 27701, 2019). Lachaud (2020) dice que este estándar es uno de los mejores métodos para poner en práctica el principio de *accountability*, ya que proporciona criterios técnicos uniformes que ayudan a implementar y certificar los sistemas de privacidad en la organización.

Como complemento crucial para el ecosistema educativo digital, la norma ISO 27018, también del 2019 establece un código de buenas prácticas para la protección de la información de identificación personal (*Personally Identifiable Information*, PII) en servicios de computación en la nube pública orientados a encargados del tratamiento. Dado que las IES recurren frecuentemente a proveedores externos de plataformas de aprendizaje (LMS), herramientas de videoconferencia y repositorios institucionales en la nube, este estándar articula controles específicos sobre la limitación de la finalidad, la salvaguarda frente a

¹⁰ Aunque la denominación técnica oficial de estas normas incluye el acrónimo conjunto ISO/IEC —al haber sido desarrolladas por la *International Organization for Standardization* (ISO) y la *International Electrotechnical Commission* (IEC) mediante su Comité Técnico Conjunto (ISO/IEC JTC 1)—. En la literatura jurídica, la práctica corporativa y la regulación sectorial es habitual referirse a ellas de manera simplificada como «normas ISO». Por lo tanto, para ahorrar en estilo y seguir el uso común en el área de *compliance* y Derecho de la tecnología, en esta investigación se usará el término general "ISO", reconociendo claramente su naturaleza técnica compartida.

tratamientos no autorizados de los datos del estudiantado, la eliminación segura y el estricto cumplimiento de las instrucciones institucionales (ISO, 2019b).

Por lo que se refiere al consentimiento informado en el entorno universitario, la norma ISO 29184 del 2020 formula directrices para el diseño de avisos de privacidad y mecanismos de obtención del consentimiento en servicios digitales. Exige que la información brindada al titular sea clara, accesible y comprensible antes de adoptar una decisión (ISO 29184, 2020). Estas pautas son cruciales en las IES, donde el uso de las herramientas digitales requiere un mayor cuidado en la transparencia, asegurando que el consentimiento que dan los usuarios sea realmente consciente e informado. En este sentido, Obar y Oeldorf-Hirsch (2020) advierten que la eficacia del consentimiento no reside en su mera formalización ritual, sino en que la información suministrada resulte efectivamente cognoscible y accesible para los destinatarios. Ahondando más en esta cuestión del consentimiento informado, la especificación técnica ISO 27560 del 2023 aporta una contribución crucial al *compliance* universitario al definir una estructura estandarizada para el registro del consentimiento (*Consent Record Information Structure*). El objetivo de la ISO es fijar el conjunto mínimo de datos que debe conservarse para probar cuándo, cómo y bajo qué condiciones específicas fue otorgado, modificado o revocado dicho consentimiento (ISO 27560, 2023).

En el ámbito de la educación, es fundamental tener registros claros y que se puedan usar en diferentes sistemas para mejorar la transparencia del consentimiento en el uso de herramientas de análisis del aprendizaje y sistemas de evaluación digital.

Finalmente, ante la creciente incorporación de la IA en la educación superior —mediante algoritmos de analítica predictiva, tutorías automatizadas, evaluación del desempeño o procesos de admisión—, la norma ISO 42001 del 2023 establece los requisitos para implantar, mantener y mejorar un Sistema de Gestión de IA (*Artificial Intelligence Management System*, AIMS). Este estándar proporciona la estructura general para la gestión de riesgos algorítmicos, la prevención de sesgos, la transparencia y la supervisión humana continua a lo largo del ciclo de vida de los sistemas de IA (ISO 42001 2023), erigiéndose como el marco organizativo dentro del cual se integran las salvaguardas de privacidad y derechos fundamentales de la comunidad universitaria.

Para sistematizar el marco técnico expuesto en este apartado, a continuación, se presenta un cuadro sintético de las normas ISO analizadas. En dicho instrumento se detalla la denominación de cada estándar, su año de publicación, el tema principal que aborda y el tipo de documento, ofreciendo una visión consolidada de su función en la gestión del cumplimiento normativo.

Tabla 4 . Principales estándares ISO para la implementación de la gobernanza de la IA y del consentimiento informado en las IES.

Estándar	Año	Tema Principal	Tipo de Documento
ISO/IEC 27701	2019	Gestión de Privacidad (PIMS)	Norma Internacional
ISO/IEC 27018	2019	Protección de datos (PII) en Nube	Código de Prácticas
ISO/IEC 29184	2020	Avisos de privacidad y consentimiento online	Directrices
ISO/IEC TS 27560	2023	Registro técnico del consentimiento	Especificación Técnica
ISO/IEC 42001	2023	Sistema de Gestión de IA (AIMS)	Norma Internacional

Fuente: *International Organization for Standardization (ISO)* y la *International Electrotechnical Commission (IEC)*. Elaboración propia.

En conjunto, estos estándares internacionales (ISO) consolidan la evolución de la protección de datos personales hacia modelos de cumplimiento basados en el *compliance* y el principio de *accountability*, en los que el cumplimiento normativo deja de concebirse como una obligación meramente declarativa para convertirse en un proceso permanente de gestión, supervisión, documentación y mejora continua. En el contexto de la IA aplicada a la educación superior, estos estándares proporcionan los mecanismos necesarios para operacionalizar principios como la transparencia, la responsabilidad, la privacidad y el consentimiento informado, permitiendo que estos se traduzcan en procesos institucionales éticos, trazables, verificables y jurídicamente sólidos. Sobre esta base resulta posible formular un modelo integrado que articule las dimensiones jurídicas, ética, tecnológica, pedagógica y organizacional del consentimiento informado, propuesta que se desarrolla en el apartado siguiente mediante El Modelo de Gobernanza para la Confianza.

4. 4. El Modelo de Gobernanza para la Confianza (MGC)

El Modelo de Gobernanza para la Confianza (MGC) constituye la propuesta original de esta investigación para la gobernanza del consentimiento informado en entornos educativos mediados por IA. Su principal aportación consiste en integrar, dentro de una misma arquitectura institucional, el marco ecológico de Chan (2023), los marcos de IA fiable y gestión de riesgos de la Unión Europea y el NIST (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023), la perspectiva pedagógica ACE del U. S. Department of Education (2023) y el modelo de la confianza de Covey (2006). Como parte de esta integración, el MGC incorpora una dimensión formativa como contribución estructural de esta investigación, al reconocer la alfabetización en IA como una función institucional diferenciada y necesaria para fortalecer la confianza, la supervisión humana y el ejercicio de un consentimiento genuinamente informado.

Aunque Chan (2023) incorpora la capacitación y el desarrollo de capacidades dentro de la dimensión operativa de su marco ecológico, el MGC propone otorgar a la formación un carácter autónomo y estructural. Esta decisión no supone desconocer el componente formativo presente en dicho marco, sino diferenciar dos finalidades institucionales distintas. Por una parte, la dimensión operativa asegura la implementación y capacitación técnica y el funcionamiento adecuado de los sistemas de IA. Por otra parte, la dimensión formativa ayuda a docentes y estudiantes a desarrollar habilidades éticas, críticas y digitales. Estas habilidades les permiten entender, supervisar, cuestionar y usar estos sistemas de manera responsable. En consecuencia, la alfabetización en IA no se concibe únicamente como apoyo para la operación tecnológica, sino como una condición para la supervisión humana significativa, la autonomía de los usuarios y el ejercicio de un consentimiento verdaderamente informado (Chan, 2023; UNESCO, 2021; U. S. Department of Education, Office of Educational Technology, 2023).

Para detallar nuestro modelo para la confianza, extrapolamos al ámbito de las IES en México el planteamiento de Covey (2006), quien sostiene que la confianza se apoya en dos dimensiones: carácter y competencia. Mientras el carácter remite a integridad, motivo e intención, la competencia se vincula con capacidades, habilidades, resultados e historia (Covey, 2006, p. 31).

En el contexto de las IES en México, el carácter incluye aspectos como la lealtad, la integridad y la responsabilidad de las instituciones hacia sus estudiantes, profesores e investigadores. Por otro lado, la competencia se refiere a la habilidad técnica, la claridad en las operaciones y el cumplimiento de las normas. Esta perspectiva es importante para el consentimiento informado: la IES necesita mostrar tanto una intención ética (carácter) como su capacidad institucional, que incluye el conocimiento sobre IA, para que el estudiante entienda cómo se tratarán sus datos y los riesgos que esto conlleva.

La falta de cualquiera de las dos dimensiones reduce la confianza, algo que Covey (2006) llama fricción institucional: una confianza baja aumenta la fricción, ya sea por acciones no éticas o por comportamientos éticos pero ineficaces, ya que las buenas intenciones casi nunca reemplazan un mal juicio.

En el MGC, la confianza no constituye una dimensión adicional ni un resultado accesorio, sino el principio articulador que da sentido a las cuatro dimensiones del modelo. Por lo tanto, la gobernanza, el funcionamiento, la mediación educativa y la capacitación institucional son importantes porque mejoran la calidad y la habilidad de la institución, permitiendo un acuerdo realmente informado.

La conexión entre la visión ecológica (Chan, 2023), la iteración de gobernanza y gestión de riesgos (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023) y la confianza sigue el principio de responsabilidad proactiva (*accountability*), que exige que se adopten medidas organizativas y técnicas para garantizar la protección efectiva de derechos y la corrección de fallas.

El MGC conserva las tres dimensiones propuestas por Chan (gobernanza, operativa y pedagógica), pero las reconfigura al incorporar una cuarta dimensión: la formativa. Esta incorporación busca reconocer la alfabetización en IA como una función importante dentro de la institución, que puede mejorar la supervisión humana, el consentimiento informado y la confianza.

La incorporación de esta nueva dimensión proviene de distinguir entre dos niveles dentro de la dimensión operativa mencionada por Chan. El técnico-operativo está a cargo de las áreas de Tecnologías de la Información (TI), responsables de garantizar seguridad, explicabilidad y privacidad. El formativo está orientado a desarrollar las competencias éticas y digitales necesarias para que docentes y estudiantes puedan supervisar, comprender y cuestionar el uso de IA. Esta decisión tiene implicaciones en el diseño del modelo: al ver la formación como un aspecto independiente, el MGC permite establecer objetivos, responsables, procesos, indicadores y formas de evaluar específicos para el desarrollo de habilidades en IA. De este modo, la alfabetización deja de concebirse como una actividad complementaria de la implementación tecnológica para convertirse en un eje permanente de la gobernanza institucional, orientado a fortalecer la supervisión humana, el pensamiento crítico y el ejercicio de un consentimiento genuinamente informado (UNESCO, 2021; U. S. Department of Education, Office of Educational Technology, 2023).

La dimensión formativa converge de manera natural con la propuesta ACE del U. S. *Department of Education* (2023) y con los marcos UE/NIST. Esto se debe a que la alfabetización y la capacitación son condiciones para la supervisión humana significativa y para el uso responsable de la IA. En consecuencia, la dimensión formativa aporta el componente pedagógico que hace operativas la transparencia, la rendición de cuentas y la

gestión iterativa de riesgos (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023).

De manera complementaria, el U. S. *Department of Education* (2023) sostiene que, cuando el profesorado permanece dentro de los procesos de decisión, el estudiantado confía más en los resultados y participa de manera más crítica en su aprendizaje.

La dimensión formativa es el proceso que la institución usa para enseñar a docentes y estudiantes las habilidades éticas y digitales que necesitan para entender, controlar y usar la IA de forma responsable. De este modo, fortalece la competencia institucional —pilar de la confianza según Covey (2006)— y crea las condiciones para un consentimiento genuinamente informado.

La necesidad de una dimensión formativa autónoma dentro del MGC encuentra respaldo empírico directo en la ENIAG 2025: pese al uso generalizado de la IA, la autopercepción de competencia en su manejo se mantiene baja en todos los tipos de institución y géneros, sin superar 5.5 puntos sobre 10, mientras que el 91% de los docentes y el 76.2% de los estudiantes de instituciones públicas manifiestan la necesidad de tomar cursos vinculados a la IA (SEP, 2026). Esta disociación entre uso intensivo y competencia real confirma que la alfabetización en IA no puede subsumirse en la dimensión operativa de Chan (2023), sino que constituye una necesidad institucional diferenciada y urgente, tal como lo plantea el MGC.

La integración orgánica de los elementos antes expuestos nos permite presentar un modelo de gobernanza basado en la confianza. En este modelo, convergen la estructura (Chan, 2023), la gobernanza iterativa y gestión de riesgos (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023), la mediación docente (U. S. *Department of Education*, 2023) y la formación institucional permanente. En este modelo consideramos que la confianza es el resultado de una gobernanza ética, una operación técnica segura —mediante las ISO— y una práctica pedagógica centrada en las personas.

La confianza se convierte en el hilo conductor que une la arquitectura institucional (Chan, 2023) con la gobernanza verificable y la gestión iterativa de riesgos de la IA (Comisión Europea, 2019; NIST, 2023).

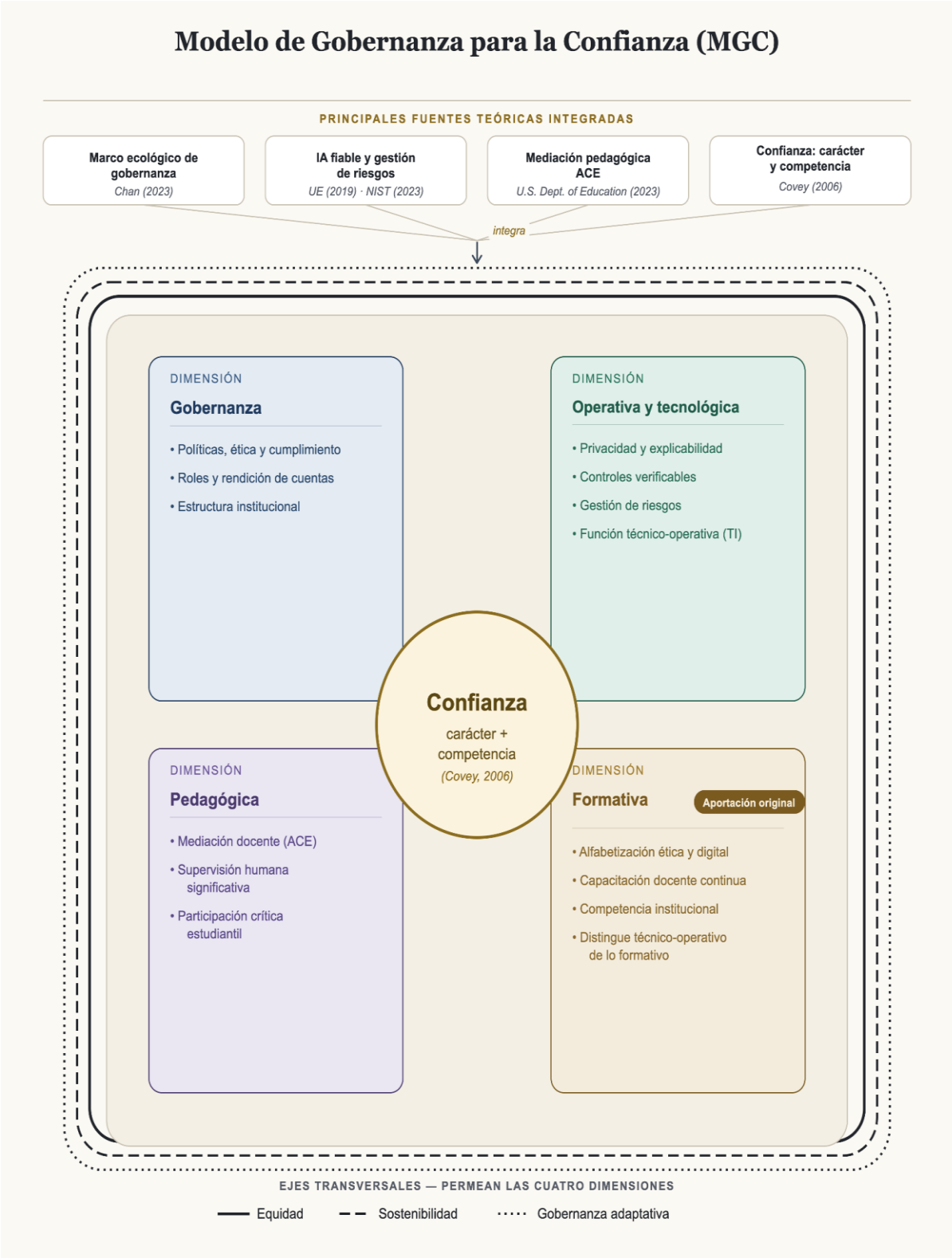
Aunado a lo anterior es importante establecer la integración de ejes transversales referidos a la equidad, la sostenibilidad y la gobernanza adaptativa. Esto obedece a que tales principios no constituyen funciones operativas aisladas —a diferencia de los cuatro pilares fundacionales del modelo— sino que son imperativos que permean íntegramente cada una de las dimensiones propuestas.

La equidad en la gobernanza de la IA es el primer elemento transversal. La OCDE (2024) preserva el “crecimiento inclusivo” como eje rector en la revisión de sus pautas sobre IA, insistiendo en que los provechos de dicha tecnología han de repartirse proporcionalmente entre los diversos sectores sociales, trascendiendo la ventaja de quienes poseen la infraestructura para una implementación temprana. Esta disparidad en la disponibilidad de recursos constituye una de las preocupaciones centrales en la producción académica sobre las universidades de la región: Kanobel (2026) señala que la brecha digital en numerosas instituciones públicas de América Latina limita sustancialmente la facultad del alumnado y del profesorado para hacer valer los derechos y salvaguardas que los instrumentos de gobernanza suelen proclamar únicamente en el plano normativo.

La sostenibilidad digital alude al impacto ambiental derivado de la infraestructura y el cómputo que subyacen a los sistemas de IA. Kanobel (2026) posiciona la «sostenibilidad digital» como una de las cinco dimensiones ineludibles para la gobernanza responsable de la IAGen en las instituciones de educación superior. A partir de su revisión de políticas institucionales, el autor documenta que esta área, junto con la explicabilidad algorítmica, presenta la brecha más crítica entre la formulación declarativa de las políticas y su implementación práctica. Su configuración como dimensión transversal —y no como un componente aislado circunscrito exclusivamente a la dimensión operativa y tecnológica— reconoce que la huella ambiental no emana de una única función institucional, como la infraestructura de TI, sino que es resultado de decisiones estratégicas en los ámbitos de investigación, docencia y gestión administrativa al seleccionar los modelos de IA a emplear y establecer su frecuencia de uso.

El último elemento transversal es la gobernanza adaptativa. Robert y McCormack (2024) evidencian que una de las carencias medulares identificadas en su consulta a novecientos especialistas en tecnología para la educación fue la inexistencia de protocolos de revisión periódica; tal omisión provocó que un alto porcentaje de universidades operara bajo marcos normativos obsoletos ante las transformaciones técnicas y legales del propio año de emisión. Kanobel (2026) coincide con este diagnóstico al reivindicar la “gobernanza adaptativa” —la potestad de un marco para ajustarse según los hallazgos de su ejecución efectiva— frente a una innovación que desborda los tiempos ordinarios de la regulación universitaria.

Figura 9. El Modelo de Gobernanza para la Confianza (MGC)



Nota. Elaboración propia con apoyo de Claude Sonnet 5 (Anthropic, 2026), mediante un proceso iterativo iniciado con el prompt "A partir de este texto, genera una infografía del Modelo de Gobernanza para la Confianza (MGC). En la parte superior incluye un recuadro con las principales fuentes teóricas integradas: el

marco ecológico de gobernanza (Chan, 2023), los marcos de IA fiable y gestión de riesgos (UE, 2019; NIST, 2023), la mediación pedagógica del modelo ACE (U.S. Department of Education, 2023) y el criterio de confianza como carácter y competencia (Covey, 2006), con una flecha que indique que estas fuentes se integran en el modelo. Debajo, organiza las cuatro dimensiones del modelo —Gobernanza, Operativa y tecnológica, Pedagógica y Formativa— en cuatro cuadrantes, cada uno con sus componentes en viñetas, y marca la dimensión Formativa como aportación original. En el centro coloca un círculo que represente la confianza (carácter + competencia, Covey, 2006), de manera que toque visualmente a las cuatro dimensiones. Al final, añade los ejes transversales que permean las cuatro dimensiones (equidad, sostenibilidad y gobernanza adaptativa), diferenciados con distintos estilos de línea."Prompt refinado mediante instrucciones sucesivas de la autora para el diseño gráfico, a partir del contenido y la estructura desarrollados por la autora.

Asimismo, la Figura sintetiza la configuración del Modelo de Gobernanza para la Confianza (MGC) como una arquitectura institucional en la que convergen (a) el marco ecológico de Chan y sus dimensiones de gobernanza, operación y pedagogía, ampliadas aquí con una dimensión formativa; (b) la gestión iterativa de riesgos y controles verificables promovidos por marcos de IA fiable y gestión del riesgo; y (c) la mediación humana significativa en los procesos educativos y decisorios (Chan, 2023; Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023; U. S. *Department of Education*, 2023). En consecuencia, el MGC no se limita a declarar principios, sino que exige mecanismos que permitan demostrar, con evidencia, el cumplimiento de obligaciones y la corrección de fallas bajo un enfoque de responsabilidad proactiva.

La implementación de este modelo puede apoyarse en estándares internacionales de cumplimiento normativo (*compliance*), particularmente las normas ISO relacionadas con la gestión de la privacidad, consentimiento informado y la gobernanza de la IA. Estas normas, como ya hemos expuesto, establecen requisitos y directrices para implementar sistemas de gestión, controles documentados y mecanismos de mejora continua que permiten operacionalizar los principios jurídicos y éticos que sustentan el modelo (Lachaud, 2020).

Una vez definido el MGC, el siguiente paso será traducir este marco conceptual en mecanismos institucionales concretos que permitan su aplicación efectiva en la vida cotidiana de las IES.

4. 5. Implementación del Modelo de Gobernanza para la Confianza (MGC)

Como habíamos señalado en el inciso 4.3.1, el diagnóstico del Observatorio Interinstitucional de IA en la Educación Superior en México (OIIAES, 2026), mostró la ausencia de trazabilidad institucional —no sólo normativa, sino operativa— de las IES mexicanas, es por esto que se considera muy pertinente el modelo que a continuación se presenta como una herramienta de gobernanza.

La utilidad del MGC depende de su capacidad para traducir los principios de gobernanza en prácticas institucionales verificables. En consecuencia, este apartado desarrolla los mecanismos mediante los cuales las cuatro dimensiones del modelo se operacionalizan en

protocolos, estructuras organizacionales, controles técnicos, prácticas pedagógicas y estrategias de formación, con el propósito de garantizar un consentimiento informado auténtico en las IES.

Para conseguir un consentimiento auténtico es necesario un enfoque sistémico que abarque todas las dimensiones del modelo, en el contexto general de la protección de datos. Su implementación requiere convertir esta arquitectura conceptual en protocolos, estructuras organizacionales, controles técnicos y actividades formativas que orienten la actuación institucional.

A continuación, se desarrolla la implementación de cada una de las cuatro dimensiones del MGC, mostrando cómo sus componentes estratégicos, operativos, pedagógicos y formativos se articulan para proteger los datos personales del estudiantado y hacer efectivo un consentimiento informado auténtico en las instituciones de educación superior.

4. 5. 1. Dimensión de Gobernanza

La dimensión de gobernanza constituye el pilar estratégico del modelo: actúa como garante de los derechos fundamentales del estudiante y establece las condiciones de posibilidad para un consentimiento libre de coerción (Chan, 2023). Para ejercer esa función, la Alta Dirección fija políticas, designa responsables y establece protocolos que implementan la protección de la privacidad y el consentimiento informado (NIST, 2023; International Organization for Standardization, 2023).

La implementación de esta dimensión puede apoyarse en estándares internacionales de gestión que proporcionan criterios verificables para documentar, supervisar y mejorar continuamente los procesos institucionales. Entre ellos destacan la ISO 27701 para la gestión de la privacidad, la ISO/IEC 42001 para la gobernanza de sistemas de IA, la ISO 29184 para los mecanismos de consentimiento y la ISO TS 27560 para el registro verificable del consentimiento. En conjunto, estos estándares funcionan como instrumentos de compliance que fortalecen la responsabilidad proactiva (*accountability*) del modelo (ISO, 2019a, 2020, 2023).

Asimismo, la creciente autonomía de los sistemas de IA, particularmente de la IA agéntica, refuerza la necesidad de construir marcos de gobernanza que aseguren que las decisiones y acciones ejecutadas por estos sistemas permanezcan alineadas con los fines educativos, los derechos fundamentales y la dignidad de las personas, especialmente en contextos donde el tratamiento de datos personales y la toma automatizada de decisiones pueden impactar directamente en estudiantes y docentes (UNESCO, 2023).

4. 5. 1. 1. Protocolos

Los protocolos permiten una respuesta sistemática y coherente ante situaciones recurrentes, elevan la calidad de la gestión institucional y reducen márgenes de error (García & Curiel, 2021).

En el MGC, los protocolos constituyen los principales instrumentos mediante los cuales la dimensión gobernanza se traduce en actuaciones institucionales concretas. En conjunto, los protocolos propuestos buscan proteger el consentimiento informado desde distintas etapas del ciclo de vida de los sistemas de IA: desde la gestión de datos y la supervisión algorítmica hasta la formación de la comunidad universitaria y la garantía de los derechos del estudiantado y la alfabetización de la comunidad universitaria.

A. Protocolo de gobernanza de datos

Este protocolo establece los criterios institucionales para el tratamiento ético, jurídico y seguro de la información (Osorio Sanabria et al., 2017). Articula finalidades, procedimientos, responsables y criterios de supervisión. Su implementación se apoya en la norma ISO 27701, la cual proporciona directrices para la gestión de la información de privacidad y la protección de datos personales.

Asimismo, integra los elementos de gobernanza iterativa y gestión de riesgos propuestos por los marcos de la Unión Europea y de Estados Unidos, estableciendo un ciclo sistemático de identificación de riesgos, diálogo entre actores educativos, evaluación de impactos e implementación de controles, con revisión periódica y rendición de cuentas (Comisión Europea, 2019; NIST, 2023).

B. Protocolo de Auditoría y Monitoreo Algorítmico

Este es un proceso sistemático de revisión técnica, ética y legal de los sistemas de IA, con el fin de identificar sesgos, errores, opacidad y riesgos que afecten la equidad, la transparencia o los derechos fundamentales (Bustelo Gracia, 2025).

La auditoría algorítmica periódica se alinea con estándares internacionales como el AI Act (2024), el cual considera obligatorias las auditorías de sistemas de alto riesgo para la educación, considerada un sector sensible. Asimismo, la implementación de este protocolo puede fortalecerse mediante la ISO 42001, que establece requisitos para la gestión responsable de sistemas de IA.

El monitoreo algorítmico es un proceso continuo de observación, recolección de datos y análisis del comportamiento de un algoritmo desplegado, a fin de asegurar que opere bajo criterios de equidad, eficiencia y legalidad; el Reglamento de IA europeo lo consagra como obligación de seguimiento posterior a la comercialización para los sistemas de alto riesgo

(Parlamento Europeo y Consejo de la Unión Europea, 2024, art. 72), y marcos como el del NIST (2023) lo estructuran como función permanente de medición y gestión.

Este protocolo establece un monitoreo sistemático de los sistemas de IA utilizados en procesos sensibles, como evaluaciones, becas y admisiones. Para ello, incorpora el análisis de sesgos, la participación estudiantil y la revisión periódica por el Comité de Ética y Gobernanza.

C. Protocolo de equidad y alternativas: el respeto al derecho de oposición

Para que el consentimiento sea verdaderamente libre, este protocolo garantiza recursos, rutas y opciones cuando el estudiante decida no consentir el uso de sistemas de IA en su proceso de enseñanza-aprendizaje. En este sentido, Crawford (2021) advierte que la autonomía resulta ilusoria si el costo de no consentir es la exclusión educativa. En consecuencia, cuando un estudiante ejerce su derecho de exclusión (*opt-out*), la IES proporciona una alternativa pedagógica equivalente en calidad.

Este protocolo refuerza la confianza en la IES al demostrar coherencia entre el discurso de la protección del consentimiento y las condiciones reales para ejercerlo por parte del estudiante, sin ver afectado su derecho a la educación.

D. Protocolo de alfabetización en IA

La alfabetización en IA se define como un conjunto de competencias que permiten a las personas evaluar críticamente las tecnologías inteligentes, comunicarse eficazmente con ellas y comprender su lógica subyacente. La alfabetización trasciende la mera capacitación operativa (saber usar la herramienta) para fomentar una comprensión sociotécnica profunda sobre cómo funcionan los algoritmos y qué efectos producen (Long y Magerko, 2020).

Desde la perspectiva del consentimiento, la alfabetización permite un consentimiento auténticamente informado. Sin una alfabetización adecuada que permita comprender los riesgos asociados a la privacidad y al sesgo algorítmico, el consentimiento otorgado por el usuario carece de los elementos necesarios para considerarse plenamente informado.

Este protocolo establece las bases para los programas de formación dirigidos tanto a docentes como a estudiantes. En este sentido, la ANUIES (2025) advierte que la inversión más crítica para las IES es la capacitación docente y la alfabetización digital orientada al desarrollo del pensamiento crítico del estudiantado. Además, se operacionaliza mediante programas de capacitación continuos y especializados, conforme a la dimensión formativa del MGC, cuyos aspectos específicos para su implementación se desarrollan en la dimensión formativa del MGC.

E. Protocolo de objeción algorítmica

Este protocolo establece las rutas institucionales para que los estudiantes puedan objetar o solicitar la revisión de decisiones total o parcialmente automatizadas que impacten su trayectoria educativa. Su finalidad es garantizar el respeto a los derechos fundamentales, la intervención humana cuando resulte necesaria y la justicia algorítmica (Floridi et al., 2018; Susser et al., 2019).

F. Protocolo de consentimiento informado digital

Las IES deben crear este protocolo que se aplicará siempre que se utilicen sistemas basados en IA que conlleven tratamiento de datos personales y que generen decisiones automatizadas para estudiantes y docentes (ANUIES, 2025, p. 66). Para su implementación, puede tomarse como referencia la norma ISO 29184 sobre avisos de privacidad y obtención del consentimiento en entornos digitales, así como la especificación técnica ISO TS 27560 para la gestión del consentimiento.

4. 5. 1. 2. Designación de responsables

La implementación efectiva de los protocolos y políticas del MGC requiere la designación de responsables con funciones claramente definidas para asegurar su coordinación, seguimiento y mejora continua.

El MGC contempla al menos dos figuras clave: (a) la persona o departamento responsable de datos personales —previsto en el artículo 29 de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP, 2025) (figura funcionalmente cercana al *Data Protection Officer*, DPO)— y (b) un comité colegiado y multidisciplinario.

Para el órgano colegiado, el MGC contempla la creación de un Comité de Ética en IA (CE-IA), cuyas funciones son la deliberación, evaluación y acompañamiento en la incorporación responsable de tecnologías automatizadas. Este comité es multidisciplinario e incluye representación docente y estudiantil como práctica de codiseño ético (*U. S. Department of Education*, 2023). Asimismo, cuenta con una rotación periódica de sus integrantes y presenta informes periódicos.

4. 5. 1. 3. Políticas

Las políticas proporcionan marcos normativos que guían la acción y la toma de decisiones, dando estabilidad y dirección en contextos de incertidumbre y, al aplicarse a situaciones similares, ofrecen una orientación clara sobre el rumbo que deben seguir las actividades institucionales de un mismo tipo (Ballart, 1993).

El MGC incorpora políticas que establecen medidas técnicas de protección de datos, como el cifrado, el control de acceso, la anonimización, la seudonimización y la seguridad de la infraestructura.

Aunque los proveedores realizan el tratamiento de los datos, la responsabilidad última recae en la IES. Por eso las políticas deben exigir que los proveedores de sistemas tecnológicos educativos que usen IA cumplan la legislación nacional e internacional aplicable. Incluir legislación internacional es relevante porque la movilidad e intercambio de estudiantes puede implicar el cumplimiento de normas de sus países de origen. El MGC establece que los sistemas adquiridos cumplan con las Normas Oficiales Mexicanas (NOM) aplicables y con estándares internacionales de seguridad de la información, como ISO/IEC 27001, que ofrece un sistema de gestión para identificar, evaluar y mitigar riesgos asociados al manejo de datos.

En cuanto a las políticas, esta propuesta se centra en dos ámbitos estratégicos: las políticas institucionales para la contratación de terceros y la alineación con las políticas públicas nacionales.

A. Política de revisión de contratos tecnológicos

Dentro de las políticas queremos destacar la de adquisición (*procurement*), ya que con ella se inicia el ciclo de datos que se recopilan de los estudiantes. Por lo mencionado, es importante crear una política de compras que requiera verificar que los contratos con proveedores externos cumplan con la protección de datos, la soberanía digital y los propósitos legítimos. Además, se sugiere establecer una política para evaluar los términos y condiciones de los sistemas de IA antes de comprarlos. Esto aseguraría la soberanía digital y el cumplimiento de las normas utilizando estándares como ISO 27001 y el marco de privacidad desde el diseño.

B. Vinculación con una propuesta de política pública nacional

Nuestra propuesta de implementación de un plan de gobernanza se alinea con estrategias a nivel nacional. Tomamos como ejemplo las recomendaciones de algunos investigadores de IES pertenecientes a la ANUIES (2025):

1. El Marco Nacional para Uso Responsable de la IA: que propone un código ético nacional y observatorios éticos de la IA en las IES (p. 31).
2. El Programa Nacional de Formación y Alfabetización de la IA (p. 190).
3. Comité Nacional de Ética de la IA en Educación, un órgano multidisciplinario encargado de desarrollar directrices éticas y evaluar casos complejos (p. 111).

El CE-IA de cada institución puede fungir como implementador del Código Nacional de Ética de la IA y primer interlocutor ante el Consejo Nacional, convirtiendo la política pública nacional en práctica institucional concreta.

4. 5. 1. 4. Registro Institucional de Sistemas de IA y Taxonomía de Casos de Uso por Nivel de Riesgo (heurística UE)

Para que la dimensión de gobernanza sea transferible y auditable, se propone crear un registro institucional de sistemas de IA que identifique, para cada herramienta o módulo (propio o de terceros), su finalidad, población afectada, datos tratados, proveedor, modelo de despliegue, periodo de retención, mecanismos de supervisión humana y ruta de atención de quejas. En consecuencia, este registro debe funcionar como fuente única de rendición de cuentas para determinar si el consentimiento es procedente, suficiente o excepcional, y para decidir qué salvaguardas deben activarse.

El registro se robustece si se acompaña de una taxonomía mínima de casos de uso y una clasificación por nivel de riesgo. Este enfoque es congruente con la lógica de gestión del riesgo del Reglamento (UE) 2024/1689 (Ley de IA), que distingue prácticas prohibidas, sistemas de alto riesgo y obligaciones diferenciadas según el impacto sobre derechos fundamentales (Parlamento Europeo y Consejo de la Unión Europea, 2024).

Por lo anterior, se propone una taxonomía mínima para IES y su clasificación en cuatro niveles —prohibido, alto riesgo, riesgo limitado y riesgo mínimo— como estrategia para decidir cuándo el consentimiento basta, cuándo es insuficiente y cuándo el despliegue debe ser excepcional o descartarse, en coherencia con la lógica de gestión del riesgo del Reglamento (UE) 2024/1689 (Parlamento Europeo y Consejo de la Unión Europea, 2024). Esta clasificación debe vincularse con una regla operativa: a mayor riesgo, mayor exigencia de documentación, evaluación de impacto y garantías procedimentales.

Además, resulta congruente exigir medidas de alfabetización en IA para quienes operan y supervisan estos sistemas, para promover un nivel suficiente de alfabetización en IA en el despliegue institucional (Parlamento Europeo y Consejo de la Unión Europea, 2024).

Retomando los resultados del primer diagnóstico de cobertura nacional sobre la adopción de la IA en las IES mexicanas (OIIAES, 2026) se observa que en materia de infraestructura, el diagnóstico muestra una adopción todavía incipiente y desigual: sólo el 24.2% de las IES ha adquirido infraestructura tecnológica dedicada a la IA y el 37.3% ha adquirido herramientas específicas, mientras que la colaboración interinstitucional —con otras IES nacionales o con organismos gubernamentales e internacionales— no supera el 32% de los casos (OIIAES, 2026). Estos datos respaldan empíricamente uno de los componentes que el MGC incorpora en su dimensión de gobernanza: el registro institucional de sistemas de IA y la necesidad de mecanismos de colaboración interinstitucional como condición para una adopción estratégica, y no meramente operativa.

4. 5. 1. 5. Doble Evaluación de Impacto *Ex Ante* y Revisión Continua: DPIA, EIA y EIDF

Cuando un caso de uso se clasifica como de alto riesgo, las IES deben exigir una doble evaluación previa a su puesta en marcha: (a) una Evaluación de Impacto en Protección de Datos (EIPD) y (b) una Evaluación de Impacto Algorítmico (EIA), complementada, cuando corresponda, por una evaluación de impacto en derechos fundamentales prevista para ciertos despliegues de alto riesgo en el marco europeo (Parlamento Europeo y Consejo de la Unión Europea, 2024).

Asimismo, la DPIA permite identificar y mitigar riesgos derivados del tratamiento de datos personales cuando este es probablemente de alto riesgo, especialmente si incluye perfilado, monitoreo sistemático o uso de categorías sensibles; en el estándar europeo, esta exigencia se asocia con el artículo 35 del Reglamento (UE) 2016/679 (GDPR) (Parlamento Europeo y Consejo de la Unión Europea, 2016).

Por su parte, la AIA debe centrarse en riesgos algorítmicos: sesgo y discriminación, errores sistemáticos, robustez, ciberseguridad, trazabilidad, y efectos sobre derechos como igualdad, debido proceso y libertad académica. En consecuencia, para operacionalizarla se sugiere un ciclo de gobernanza inspirado en el AI Risk Management Framework del NIST con evidencias documentales que faciliten auditoría interna y externa (2023).

Finalmente, ambas evaluaciones deben traducirse en un plan verificable de mitigación (controles técnicos, controles organizacionales, responsables, plazos y criterios de aceptación del riesgo), así como en una revisión periódica basada en incidentes, quejas, sesgos detectados y cambios en proveedores, datos o modelos.

Para facilitar la adopción institucional del registro, taxonomía y doble evaluación, se incluye la siguiente tabla:

Tabla 5. Taxonomía mínima de casos de uso × nivel de riesgo × salvaguardas.

Caso de uso	Nivel de riesgo	Salvaguardas mínimas	¿Opt-out?	¿Revisión humana?	¿DPIA/AIA?
Admisión/becas (perfilamiento)	Alto	Transparencia reforzada; no automatización plena; auditabilidad	Sí	Sí	Sí
Evaluación/calificación	Alto	Explicabilidad; revisión humana; mecanismo de contradicción	Sí	Sí	Sí
Proctoring/vigilancia remota	Alto	Proporcionalidad; minimización; alternativas equivalentes	Sí	Sí	Sí
Inferencia de emociones	Prohibido	No desplegar salvo excepciones estrictas; control reforzado	N/A	N/A	Sí
Analítica predictiva	Alto	Validación; control de sesgo; no estigmatización	Sí	Sí	Sí
Recomendadores	Limitado	Transparencia; control de preferencias; monitoreo	Sí	Parcial	Según caso
Chatbots/asistentes	Limitado	Aviso claro de IA; escalamiento a humano; logs	Sí	Parcial	Según caso
Apoyos de baja incidencia	Mínimo	Información básica; seguridad; minimización	Sí	No	No
Nota. Elaboración propia con base en la heurística de riesgo descrita en el texto.					

Participación ampliada de la comunidad.

La Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2024) ha subrayado recientemente la imperiosa necesidad de un “crecimiento inclusivo” en sus principios rectores, enfatizando que la arquitectura y el empleo de sistemas inteligentes deben integrar las perspectivas de los grupos impactados, trascendiendo la visión de proveedores o adquirentes. En la esfera de la educación universitaria, Robert y McCormack (2024) advierten que los modelos de gobernanza más robustos surgen de procesos de construcción colectiva —con participación de académicos, especialistas técnicos y el estudiantado—, distanciándose de políticas verticales dictadas unilateralmente. Por su parte, Kanobel (2026) reafirma esta postura al situar la “participación democrática” como un pilar indispensable para mitigar la fractura entre las intenciones declaradas en la normativa institucional y su ejecución real en los campus.

4. 5. 2. Dimensión Operativa

En la dimensión operativa se materializa la implementación de los protocolos y políticas de gobernanza de datos. En consecuencia, el área de Transformación Digital traduce los acuerdos de gobernanza en controles técnicos, infraestructura segura y mecanismos para mitigar riesgos algorítmicos y proteger el consentimiento del estudiantado.

Dentro de estos mecanismos, destacamos la privacidad desde el diseño y por defecto, así como la interfaz de transparencia, el consentimiento informado reforzado en capas (*layered*) y consentimiento dinámico¹¹.

El área responsable de operación (Transformación Digital, TD) debe llevar un registro de actividades de tratamiento, detallando los procesos que involucran datos personales. Es recomendable realizar evaluaciones de impacto documentadas y actualizadas regularmente. Finalmente, se sugiere efectuar auditorías periódicas para verificar que los proveedores cumplan con los requisitos establecidos.

En definitiva, en esta dimensión recae la aplicación de las políticas operativas definidas por la dimensión de gobernanza. La dimensión operativa traduce los mandatos de la gobernanza en controles técnicos tangibles.

4. 5. 2. 1. Evaluación de proveedores

El *National Institute of Standards and Technology* (NIST, 2023), en su Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF 1.0), establece que la función de “Gobernar” (Govern) —una de las cuatro funciones núcleo del marco— exige extender la gestión de riesgos a terceros, incluidos los proveedores externos de sistemas de IA, dado que la responsabilidad institucional no se transfiere por el hecho de subcontratar la tecnología. En la misma línea, el Grupo de Expertos de Alto Nivel en Inteligencia Artificial de la Comisión Europea (2019) incorpora la “auditabilidad” como uno de los siete requisitos de una IA fiable, la cual solo es verificable si la institución exige a sus proveedores evidencia documentada sobre el origen de los datos de entrenamiento, las medidas de mitigación de sesgo y los mecanismos de supervisión humana antes de adoptar la herramienta.

4. 5. 2. 2. Inventario y clasificación de usos por nivel

La categorización de los sistemas y aplicaciones de IA en función de su peligrosidad potencial representa el proceso indispensable para que los postulados éticos abstractos devengan en normas institucionales aplicables. En esta dirección, el NIST (2023) articula su propuesta mediante las dimensiones de “mapeo” y “medición”, las cuales exigen deslindar los escenarios de implantación y ponderar el alcance de sus efectos antes de determinar el grado de control humano indispensable. De forma análoga, el Grupo de Expertos de Alto Nivel de la Comisión Europea (2019) supedita la intensidad de sus siete exigencias sustantivas —tales como la robustez técnica, la confidencialidad y la transparencia— a la

¹¹ Otras tecnologías que protegen datos incluyen: el aprendizaje federado (Kairouz et al., 2021), que evita extraer información del dispositivo del estudiante; la IA explicable (Doshi-Velez & Kim, 2017), que permite auditar decisiones y reducir opacidad; la minimización y el cifrado (RIPD, 2019), que protegen los datos en tránsito y en reposo; y, por último, el blockchain y los contratos inteligentes, que pueden aportar trazabilidad y verificabilidad sobre acceso y consentimiento.

escala del impacto proyectado sobre los derechos individuales. Georgieva et al. (2025), al formular las pautas éticas de EDUCAUSE, sitúan la “ponderación de riesgos y beneficios” como un eje rector para la educación superior, advirtiendo que esta fiscalización debe ser ininterrumpida y no limitarse a un análisis preliminar, a fin de validar que la utilidad pedagógica prevalezca sobre los riesgos para grupos vulnerables. No obstante, la carencia de un catálogo oficial que jerarquice tecnológicamente cada solución impide la aplicación coherente de estos instrumentos, consolidando la denominada “IA en la sombra” o despliegue de herramientas al margen de la gobernanza universitaria.

4. 5. 2. 3. Configuración de Privacidad por Defecto (*technical default*)

La congruencia del sistema depende de que la tecnología no traicione la promesa de la IES y, por lo tanto, la confianza. Se ha considerado, siguiendo a Cavoukian (2012), que la IES debe garantizar que la privacidad sea por defecto, integrada desde el diseño y reforzada durante todo el ciclo de vida de los sistemas de IA. Esta privacidad por diseño responde a las preocupaciones de vigilancia algorítmica identificadas por O’Neil (2017) y Lanier (2018), y coincide con las recomendaciones de la RIPD (2019).

El área de TD configura las licencias institucionales de tipo empresarial (*enterprise*) para desactivar por defecto el entrenamiento de modelos (*model training*) con datos del usuario. En consecuencia, se implementan registros (*logs*) de auditoría que permitan verificar que los datos del estudiantado no han salido del entorno seguro, proporcionando rendición de cuentas (*accountability*), conforme a estándares como la norma IEEE 7000-2021, orientada a abordar preocupaciones éticas durante el diseño de sistemas.

4. 5. 2. 4. Interfaz de Transparencia

Como se ha señalado en múltiples ocasiones, los avisos legales extensos de privacidad resultan ineficaces para la comprensión real y, por lo tanto, para que el consentimiento esté verdaderamente informado. En consecuencia, una práctica concreta consiste en traducir en lenguaje claro y visual la información sobre dónde se almacenan los datos del estudiante, para qué se utilizarán y qué riesgos están asociados a cada herramienta.

Para lograr lo anterior, se propone la implementación de un tablero (*dashboard*) de privacidad simplificado. Este panel mostrará al estudiante, mediante un semáforo visual, el nivel de exposición de sus datos en cada herramienta activa (verde: procesamiento local de datos; amarillo: nube privada; rojo: nube pública o entrenamiento activo). Así el consentimiento del estudiante podrá ser realmente informado.

4. 5. 2. 5. Consentimiento Informado Reforzado en Capas (layered) y Consentimiento Dinámico

Para evitar que el consentimiento se reduzca a un trámite documental, se propone adoptar un modelo operativo de información en capas (*layered*) que haga *posible* comprender, decidir y revocar de manera efectiva. En consecuencia, este modelo traduce el estándar europeo de consentimiento —claridad, especificidad, ausencia de ambigüedad y posibilidad real de retirada— en elementos de interfaz y de gobernanza verificables (European Data Protection Board [EDPB], 2020).

El modelo se estructura en cuatro capas complementarias. En primer lugar, aviso breve: identifica qué sistema de IA se usa, para qué finalidad inmediata, qué datos clave intervienen y cuál es la opción real (aceptar, rechazar o usar una alternativa equivalente). En segundo lugar, aviso ampliado: detalla finalidades primarias y secundarias, transferencias a terceros, criterios de retención, medidas de seguridad y canales de ejercicio de derechos. En tercer lugar, ficha técnica del sistema de IA: describe función y límites, proveedor o desarrollador, datos de entrenamiento o ajuste cuando sea pertinente, sesgos conocidos, métricas de desempeño, supervisión humana y medidas de mitigación. Finalmente, panel de preferencias: permite granularidad por finalidad (docencia, analítica, investigación), revocación sencilla, trazabilidad de cambios y evidencia de que la revocación tiene efectos reales en los flujos de datos.

Tabla 6. Estructura del consentimiento informado por capas (contenido mínimo y evidencia).

Capa	Objetivo	Contenido mínimo	Canal/soporte	Evidencia de entrega
1. Aviso breve	Decisión rápida sin coerción	Herramienta/IA; finalidad inmediata; datos clave; alternativa sin IA; botones Aceptar/Rechazar/Alternativa	Pop-up/LMS/app institucional	Registro de visualización + elección (<i>timestamp</i>)
2. Aviso ampliado	Comprensión suficiente	Finalidades primarias/secundarias; transferencias; retención; seguridad; derechos; contacto DPO	Página desplegable + PDF accesible	Versión del aviso + acuse digital
3. Ficha técnica	Transparencia técnica comprensible	Proveedor; despliegue (local/nube); <i>training</i> on/off; sesgos conocidos; supervisión humana; métricas; límites	Ficha en Registro IA + enlace desde LMS	Control de versión + auditoría de cambios
4. Panel de preferencias	Granularidad y revocación efectiva	Permisos por finalidad; revocación 1 clic; historial de cambios; exportación de datos; opt-out	Dashboard de privacidad	Bitácora de cambios + confirmación al usuario
Fuente: Elaboración propia				

Este enfoque puede transformarse en un consentimiento dinámico, que se ve como un proceso continuo donde los estudiantes pueden modificar sus permisos con el tiempo y recibir alertas cuando cambien los propósitos, proveedores o riesgos. Aunque este concepto viene de discusiones sobre investigación con datos, su lógica se puede aplicar al entorno universitario digital porque mejora la autonomía y la transparencia durante todo el ciclo de vida de los datos (Kaye et al., 2015).

Finalmente, cuando exista un desequilibrio estructural entre la IES y el estudiantado —por ejemplo, cuando el uso de una plataforma sea condición de acceso a un servicio educativo— debe explicitar qué alternativas existen y qué salvaguardas se implementan para asegurar que el consentimiento sea libre y no meramente formal (EDPB, 2020).

4. 5. 2. 6. Evaluaciones de impacto.

Robert y McCormack (2024) identifican la gestión de datos y la evaluación continua de implementaciones de IA como parte del dominio Operaciones de su Plan de Acción, extendiendo la exigencia de gobernanza más allá del aula hacia los procesos administrativos e institucionales. Por su parte, el NIST (2023) ratifica esta tesis al disponer que el componente “Medición” de su arquitectura de control debe alcanzar a todo entorno que determine el estatus de los individuos, con independencia de su naturaleza pedagógica o de servicios.

4. 5. 3. Dimensión Pedagógica

La permanencia de los docentes en el centro del proceso de decisión, en lugar de ser sustituidos por la IA, es un factor clave para fomentar la confianza del estudiantado y su participación crítica en el aprendizaje (U. S. *Department of Education*, 2023).

En las últimas décadas hemos visto cómo el papel del docente se transforma, dejando de ser el mero poseedor del conocimiento para enfocarse en el desarrollo de competencias en sus estudiantes. Con la introducción de la IA en el proceso de enseñanza-aprendizaje, se destaca otro rol del docente que llamaremos “diseñador instruccional con IA”¹². El docente proyecta la arquitectura del aprendizaje y selecciona los sistemas de IA que pueden serle útiles en las cuatro fases esenciales del proceso: diseño, desarrollo, impartición y evaluación.

La IA puede complementar y potenciar la labor docente en cada una de esas fases, optimizando la creación de experiencias de aprendizaje, automatizando tareas técnicas y permitiendo la personalización masiva de contenidos. Sin embargo, es importante que se entienda como un complemento y no como un sustituto. Toda decisión crucial generada por IA debe ser validada y ratificada por un profesional humano cualificado. De este modo, la

¹² Este concepto proviene del diseño instruccional que es un proceso que consiste en elaborar de forma estratégica, planificada y estructurada los diferentes elementos de un curso académico o formativo, tales como temas, contenidos, actividades, recursos de apoyo y evaluaciones.

confianza del estudiante es el resultado activo derivado de la corresponsabilidad pedagógica del docente.

En este punto, el Triángulo de la Irreductibilidad Humana opera como criterio de diseño del MGC: la mediación pedagógica debe asegurar que la IA no reemplace, (a) el vínculo y la corresponsabilidad educativa (empatía), (b) la deliberación prudencial y la imputabilidad de las decisiones (juicio ético) y (c) la exploración creativa y el pensamiento crítico (pensamiento divergente). Por ello, las acciones pedagógicas siguientes se conciben como salvaguardas formativas verificables.

Para la implementación concreta de esta dimensión, el modelo propone dos acciones pedagógicas clave, la transparencia didáctica y las pausas pedagógicas.

4. 5. 3. 1. Transparencia Didáctica

La primera acción educativa es establecer una transparencia clara en la enseñanza, donde los profesores informan a los estudiantes sobre los principios pedagógicos del uso de IA y los objetivos específicos del manejo de datos. (véase Tabla 7)

El objetivo de la transparencia didáctica es situar el uso de la IA en el proceso educativo, ayudando al estudiante a entender cómo esta herramienta afecta su autonomía, su aprendizaje y su evaluación. En consecuencia, esta práctica convierte el consentimiento en un proceso concreto y dinámico, siguiendo el enfoque *ACE* (U. S. *Department of Education*, 2023), donde el docente es formador y mediador. La transparencia en la enseñanza permite un consentimiento constante, actualizado y consciente, de acuerdo con una lógica de mejora continua y gestión de riesgos que exige revisar finalidades, controles y efectos durante el ciclo de vida del sistema (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; *NIST*, 2023).

Para estandarizar la transparencia didáctica, se propone una plantilla breve para incorporar en el programa de la asignatura, dando un tiempo para que los estudiantes la puedan consultar y decidir, dando paso a un auténtico consentimiento informado. Como señala Sánchez Díaz “no basta con que el usuario consienta; debe comprender” (2023, p. 191).

La plantilla propuesta incorpora, además, un campo específico para declarar qué capacidades humanas se busca preservar y ejercitar (empatía, juicio ético y pensamiento divergente), así como los riesgos de sustitución formativa asociados al caso de uso.

Tabla 7. Transparencia didáctica

Elemento	Contenido mínimo	Ejemplo breve
Sistema (s) de IA	Nombre y finalidad	Se usará X para retroalimentación formativa.
Reglas de uso en la materia	Qué está permitido/prohibido	No se permite entregar texto generado sin atribución.
Datos	Datos tratados y minimización	No se subirán datos sensibles ni listas nominales.
Transparencia	Aviso de IA y límites	La herramienta puede equivocarse; se verificará.
Evaluación	Cómo se evalúa el uso de IA	Se evalúa el proceso y la reflexión, no solo el resultado.
Alternativas	Opción equivalente sin IA	Opción B: actividad manual con rúbrica equivalente.
Consentimiento/opt-out	Cómo ejercer control	Panel de preferencias y solicitud sin penalización.
Queja/objeción	Canales y plazos	Correo DPO; revisión humana en 10 días.
Triángulo de Irreductibilidad Humana (capacidades humanas)	Qué capacidades se preservan y cómo; riesgos de sustitución formativa; mediación docente prevista.	Empatía: coevaluación y trabajo colaborativo. Juicio ético: análisis de dilemas/sesgos y revisión humana. Pensamiento divergente: al menos 3 alternativas justificadas.
<i>Fuente: Elaboración propia</i>		

4. 5. 3. 2. Pausas Pedagógicas

Dentro de la sesión docente se propone incorporar pausas pedagógicas de reflexión crítica en actividades donde la IA interviene en la evaluación, la retroalimentación o la toma de decisiones. Estas pausas permiten discutir colectivamente los sesgos, limitaciones y supuestos de la herramienta, así como los riesgos asociados a la automatización.

Las pausas operan como un mecanismo de alfabetización ética en tiempo real, fortaleciendo la capacidad del estudiantado para identificar opacidades, interpretar resultados algorítmicos y ejercer su juicio crítico. Esta práctica se alinea con la mejora continua y la gestión iterativa de riesgos: observar el contexto de uso, valorar impactos, ajustar controles y retroalimentar a la gobernanza conforme se identifican nuevos riesgos (NIST, 2023; Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019).

Para operacionalizar estas pausas bajo el criterio del Triángulo de la Irreductibilidad Humana, se sugiere una “pausa triangular” breve (5–8 minutos) que active, de manera explícita, las tres capacidades formativas:

- Pausa empática (relacional): ¿quién podría verse afectado por el uso de la herramienta (p. ej., estudiantes con trayectorias o necesidades diversas) y de qué manera?

- Pausa prudencial (juicio ético): ¿qué supuestos y criterios de decisión incorpora la herramienta? ¿Qué riesgos (sesgo, error, vigilancia, dependencia) deben mitigarse y quién responde por ellos?
- Pausa creativa (pensamiento divergente): ¿qué alternativas equivalentes existen (sin IA, con otra herramienta, o con un método tradicional) y cómo se justificaría la elección en términos de aprendizaje y sentido?
- Esta pausa triangular refuerza que el consentimiento se forma en el aula y que la confianza institucional depende de prácticas pedagógicas que sostienen el vínculo humano, la deliberación responsable y la creatividad crítica, incluso cuando la IA participa en evaluación o retroalimentación.

Finalmente, las pausas consolidan la idea central del capítulo: el consentimiento se forma. El aula es un espacio de análisis crítico y no de aceptación acrítica de la tecnología. Estas prácticas trasladan al aula los principios ACE (U. S. *Department of Education*, 2023) y la lógica iterativa de Comisión Europea, 2019 y NIST (2023), garantizando al docente como mediador y asegurando que la práctica se revise, evalúe y ajuste continuamente. Así, el aula se consolida como el espacio donde la confianza institucional se experimenta cotidianamente.

Para lograr lo anterior, es necesario que el docente esté bien preparado para ser un mediador en el uso ético, práctico y didáctico de la IA. Por esta razón, esta nueva dimensión que proponemos, como contribución propia, es fundamental.

4. 5. 3. 3. Autoría Responsable.

Miao y Holmes (2023), en las directrices de la UNESCO relativas a la IA generativa en el ámbito de la investigación, enfatizan que la utilización opaca de estas herramientas en la generación de conocimiento vulnera la integridad científica y la credibilidad social de la academia. Frente a ello, proponen que las instituciones adopten marcos de transparencia rigurosos sobre el origen de los contenidos. Kanobel (2026) profundiza en esta problemática mediante la noción de “autoría aumentada”, entendiéndola como un proceso de colaboración deliberada entre el investigador y el sistema algorítmico, fundamentado en la trazabilidad. Según este autor, las políticas universitarias en América Latina que oscilan entre el veto absoluto y la aceptación pasiva provocan una fractura entre la norma institucional y el desempeño efectivo de académicos y estudiantes. La inclusión de la investigación científica como eje diferenciado en la dimensión pedagógica faculta al MGC para regular la producción intelectual con criterios de integridad específicos, evitando que un área tan crítica para la vida universitaria permanezca bajo una normativa meramente implícita.

4. 5. 4. Dimensión Formativa o de Alfabetización

Esta dimensión de formación constituye el componente habilitador del Modelo de Gobernanza para la Confianza (MGC). Sin ella, la gobernanza, la infraestructura y la mediación pedagógica quedarían incompletas.

La UNESCO (2022) establece que la alfabetización mediática e informacional es un prerequisite indispensable para que la ciudadanía pueda ejercer control efectivo y ético sobre los sistemas de IA. Al adoptar este enfoque, se reconoce que el objetivo final es garantizar la agencia humana y la capacidad de tomar decisiones informadas en una sociedad algorítmica (Floridi et al., 2018).

Para establecer un protocolo de alfabetización en IA en la educación, es esencial distinguir entre las necesidades de conocimiento y ética de dos actores: los docentes (responsables y guías) y los estudiantes (titulares de datos y futuros ciudadanos digitales). Para implementar el protocolo de alfabetización en IA, se sugieren tres aspectos: (a) un plan transversal de alfabetización; (b) la incorporación de formación ética; y (c) la priorización de la alfabetización docente.

4. 5. 4. 1. El Plan Transversal de Alfabetización en IA

La aplicación concreta de un protocolo de alfabetización en IA tiene como primera concreción un plan transversal de alfabetización en IA, el cual debe incorporarse a los programas formativos de cualquier disciplina. Estos planes deben ser flexibles, abiertos, inclusivos y en continua evolución (Flores-Vivar & García-Peñalvo, 2023), y combinar enfoques específicos e interdisciplinarios, considerando el impacto de la IA en todas las áreas del conocimiento: ciencias, humanidades y artes. Pueden servir como inspiración los marcos de desarrollo de competencias que ha publicado la UNESCO (2024, 2024b) que centran la formación en el ser humano y en un marco ético.

En el ámbito nacional ANUIES (2025) sugiere que las IES implementen programas obligatorios y transversales de alfabetización en IA (*AI literacy*) para toda la comunidad, en consonancia con la Recomendación de la UNESCO (2021) sobre el desarrollo de capacidades.

Para concretar y dar visibilidad a este plan de alfabetización se sugiere que las IES creen micro-credenciales sobre temas relacionados con la IA. De este modo, la capacitación será continua y se organizará en módulos que certifiquen habilidades específicas. A continuación, se propone una plantilla de plan formativo por micro-credenciales para facilitar su implementación y evaluación.

Tabla 8. Plantilla de plan transversal de alfabetización en IA por micro-credenciales.

Micro-credencial	Público objetivo	Competencias	Duración	Evidencia de logro	Periodicidad
Fundamentos de IA en educación	Docentes y estudiantes	Qué es IA/GenAI; límites; riesgos	2-4 h	Cuestionario + actividad	Anual
Privacidad y datos	Docentes y administrativos	Minimización; consentimiento; <i>DPIA</i>	2-4 h	Caso práctico	Semestral
Sesgo y equidad	Docentes	Detección de sesgo; no discriminación	2-4 h	Rúbrica de análisis	Semestral
Uso pedagógico responsable	Docentes	Diseño de tareas; evaluación; integridad	3-6 h	Secuencia didáctica	Anual
Derechos y debido proceso	Estudiantes	Objeción; <i>opt-out</i> ; apelación	1-2 h	Simulación de solicitud	Anual
Seguridad y buenas prácticas	Todos	Higiene digital; <i>prompts</i> seguros; <i>phishing</i>	1-2 h	Lista de verificación	Trimestral
Fuente: Elaboración propia					

También puede complementar el plan de formación ofrecer talleres o laboratorios, así como sesiones prácticas donde docentes y estudiantes analicen juntos, por ejemplo, los términos y condiciones de una herramienta popular, traduciendo el lenguaje legal a implicaciones reales.

Por su parte, para enriquecer el plan de formación es conveniente realizar grupos focales con docentes y estudiantes, a fin de validar usos y percepciones de la IA en la IES.

4. 5. 4. 2. Dar Prioridad a la Alfabetización Docente

Dentro del MGC, la alfabetización docente debe considerarse un prerrequisito para el ejercicio de una docencia de calidad. Como señalan Flores-Vivar y García-Peñalvo (2023), lo más importante radica en formar al formador, porque así se establecen las bases para una buena docencia.

El profesorado actúa como multiplicador de la cultura ética institucional y garante de la alfabetización de los estudiantes, en su práctica docente diaria. Como señala la UNESCO (2021), empoderar a los docentes con competencias digitales avanzadas es fundamental para garantizar que la IA se utilice bajo un enfoque centrado en el ser humano. Esto asegura que la tecnología amplifique las capacidades pedagógicas sin socavar la agencia y los derechos fundamentales de los estudiantes.

La formación habilita las competencias necesarias; la dimensión pedagógica es el espacio donde dichas competencias se ejercen cotidianamente en el aula. En consecuencia, la formación transforma el consentimiento tácito en una elección consciente y ofrece al

estudiantado herramientas para analizar el uso de su facultad de consentir y comprender las consecuencias del tratamiento algorítmico en su educación.

Desde esta perspectiva, el consentimiento informado deja de entenderse como un requisito aislado o meramente administrativo, y se concibe como el resultado de un entramado institucional, pedagógico y ético que articula gobernanza, formación y práctica docente.

4. 5. 4. 3. Formación Ética

La alfabetización debe incorporar una formación ética de la IA. La formación ética en la alfabetización en IA es clave para desarrollar pensamiento crítico e interactuar con la IA. Dignum (2019) sostiene que los sistemas de IA no son herramientas neutrales, sino que incorporan decisiones morales en su código.

Entender principios como la centralidad de la persona, la justicia algorítmica y la no maleficencia capacita a docentes y estudiantes para cuestionar a los sistemas de IA, más allá de la mera “alucinación”. Es necesario conocer, comprender y discernir los dilemas éticos subyacentes en el uso de la IA, como puede ser la tensión entre la comodidad de copiar y pegar ideas sintéticas y el desarrollo de pensamiento original.

Desde la óptica específica del consentimiento, la ética es clave para que el acto de aceptación del estudiante sea consciente, en un sentido más profundo. La comprensión de los sistemas de IA requerida va más allá de saber cómo funciona una herramienta y para qué resulta útil, e incorpora la pregunta por sus implicaciones en términos de valores y derechos.

Esta dimensión de formación axiológica dota al consentimiento de una validez sustantiva, ya que, como argumentan Susser et al. (2019), cuando las personas no reconocen las formas en que las tecnologías digitales pueden influir de manera encubierta sobre sus decisiones, su autonomía se ve comprometida y disminuye su capacidad para ejercer un control consciente sobre el uso de sus datos y sobre las decisiones mediadas por sistemas de IA.

Una alfabetización sin ética puede fomentar una “IAutopía” ingenua, en la que el usuario acepta la tecnología sin cuestionar su diseño y los efectos de sus interacciones.

4. 6. Operacionalización del MGC

Con el fin de que la confianza no permanezca como un valor aspiracional, sino como un resultado verificable del MGC, se propone una matriz de operacionalización que traduce sus dimensiones en criterios observables, evidencias institucionales y métricas de seguimiento.

La matriz de operacionalización se organiza conforme a las cuatro dimensiones del MGC y la relación con el componente de confianza y establece indicadores y evidencias.

Tabla 9. Matriz de operacionalización del MGC.

Dimensión MGC	Componente de confianza	Criterio observable (indicador)	Evidencia / fuente
Gobernanza	Carácter	Funcionamiento del CE-IA y responsable/DPO con atribuciones claras	Actas; nombramientos; calendario; quórum; acuerdos
Gobernanza	Carácter	Políticas institucionales vigentes y comunicadas (IA, datos, <i>procurement</i>)	Políticas publicadas; control de versiones; comunicados
Gobernanza	Competencia	Clasificación de riesgo y evaluación ex ante de casos de uso	Inventario de casos; matriz de riesgo; dictámenes
Operativa	Competencia	Privacidad por defecto y por diseño en herramientas autorizadas	Configuración enterprise; <i>checklist</i> ; validación técnica
Operativa	Competencia	Auditoría y monitoreo algorítmico documentados (sesgo, drift, opacidad)	Informes de auditoría; bitácoras; métricas de desempeño
Operativa	Competencia	Interfaz de transparencia <i>accesible</i> para el estudiantado	<i>Dashboard</i> ; fichas por herramienta; analíticas de consulta
Formativa	Competencia	Alfabetización en IA <i>evaluable</i> para docentes (mínimo institucional)	<i>LMS</i> ; constancias; evaluaciones pre/post
Formativa	Competencia	Alfabetización en IA <i>evaluable</i> para estudiantes (mínimo institucional)	<i>LMS</i> ; constancias; evaluaciones pre/post
Pedagógica	Carácter	Transparencia didáctica sobre uso de IA/datos en cada curso	Sílabos; guías; comunicación en <i>LMS</i>
Resultado	Carácter + Competencia	Confianza institucional y consentimiento informado	Encuestas; grupos focales; reportes; observatorio

Fuente: Elaboración propia

Se propone que para la aplicación del Modelo de Gobernanza para la Confianza (MGC) se genere, además, un cuadro o tablero de mando de KPI. El propósito de este instrumento es traducir los principios del modelo en métricas operativas de seguimiento que permitan verificar su implementación y sostener un ciclo de mejora continua. Los KPI se organizan por dimensión definiendo para cada indicador su métrica/fórmula, periodicidad, responsable institucional y meta sugerida. En la dimensión de gobernanza se añade los conceptos de control y reclamación¹³.

¹³ En dicho instrumento es importante clarificar que el control se refiere a la capacidad efectiva de las personas usuarias (estudiantes/docentes) para decidir y controlar el uso de IA y de sus datos en contextos académicos, sin coerción ni penalización; por reclamación nos referimos a la capacidad de impugnar, cuestionar o apelar decisiones o resultados asociados a IA (o a procesos automatizados).

Tabla 10. Tablero de mando de KPI para la aplicación del MGC.

Dimensión	KPI	Fórmula/unidad	Frecuencia	Responsable principal	Meta sugerida
Gobernanza	G1 - Inventario de IA actualizado	(# herramientas/casos registrados ÷ # detectados) x100	Mensual	TD + CE-IA	≥95%
Gobernanza	G2 - Clasificación de riesgo completada	(# clasificados ÷ total inventario) x100	Trimestral	CE-IA	100%
Gobernanza	G3 - IA/DPIA vigente en alto riesgo	(# alto riesgo con IA/DPIA ≤12m ÷ total alto riesgo) x100	Trimestral	DPO + CE-IA	≥90%
Gobernanza	G4 - Cumplimiento del plan de auditorías	(# auditorías realizadas ÷ # planificadas) x100	Semestral	CE-IA + TD	≥90%
Gobernanza	G5 - Sesiones CE-IA realizadas	(# sesiones realizadas ÷ # planificadas) x100	Trimestral	Secretaría técnica CE-IA	≥90%
Gobernanza	G6 - Compras de IA con evaluación	(# adquisiciones de IA con evaluación previa ÷ total adquisiciones IA) x100	Por evento	Compras + DPO + TD	100%
Gobernanza	G7 - Acuerdos CE-IA cerrados con evidencia	(# acuerdos cerrados con evidencia ÷ acuerdos del periodo) x100	Trimestral	CE-IA	≥80%
Gobernanza	G8 - Políticas críticas actualizadas en ≤12 meses	(# políticas críticas actualizadas en ≤12 meses ÷ total políticas críticas) x100	Semestral	DPO + Jurídica + CE-IA	100%
Gobernanza	G9 - Ejercicio efectivo de control por usuarios	(# solicitudes de modificación/retiro de consentimiento atendidas sin penalización ÷ total de solicitudes recibidas) x100	Trimestral	DPO + CE-IA	100%
Gobernanza	G10 - Reclamaciones resueltas sobre decisiones de IA	(# reclamaciones o impugnaciones resueltas ÷ total de reclamaciones recibidas) x100	Trimestral	CE-IA + Jurídica	≥90%
Operativa	O1 - Privacidad por defecto	(# herramientas con training OFF ÷ total con opción) x100	Mensual	TD	100%
Operativa	O2 - Fichas de transparencia publicadas	(# herramientas con ficha ÷ total) x100	Trimestral	TD + DPO	≥95%
Operativa	O3 - MTTR incidentes IA/datos	Promedio (días apertura-cierre) por severidad	Mensual	TD + DPO	Críticos ≤15; Altas ≤30
Formativa	F1 - Cobertura AI literacy docente	(# docentes aprobados ÷ total docentes) x100	Semestral	Centro formación docente	≥80%
Formativa	F2 - Cobertura AI literacy estudiante	(# estudiantes aprobados ÷ total estudiantes) x100	Semestral	Centro formación + Vida estudiantil	≥70%
Formativa	F3 - Mejora pre/post en comprensión	Promedio/promedio (0-100)	Semestral	Evaluación institucional	≥ +15
Pedagógica	P1 - Cursos con declaración IA/datos	(# cursos con declaración ÷ # cursos que usan IA) x100	Semestral	Coordinaciones + docentes	≥80%
Pedagógica	P2 - Pausas pedagógicas implementadas	Promedio # pausas por curso/periodo	Semestral	Docentes	≥2/curso
Pedagógica	P3 - Revisión humana en decisiones de alto impacto	(# decisiones impacto revisadas ÷ total) x100	Trimestral	CE-IA + área responsable	100%
Resultado	R1 - Índice de Confianza en IA (0-100)	Promedio encuesta (0-100)	Semestral	Evaluación institucional / Observatorio	≥80
Resultado	R2 - Índice de consentimiento informado (0-100)	Promedio encuesta (0-100)	Semestral	DPO + Evaluación institucional / Observatorio	≥85

Nota. Elaboración propia con apoyo de Claude Sonnet 5 (Anthropic, 2026), a partir del prompt "Genera un tablero de mando de KPI para la aplicación del Modelo de Gobernanza para la Confianza (MGC). El instrumento debe traducir los principios del modelo en métricas operativas de seguimiento que permitan

verificar su implementación y sostener un ciclo de mejora continua. Organiza los indicadores por dimensión —Gobernanza, Operativa, Formativa, Pedagógica y Resultado—, y para cada uno define: la métrica o fórmula de cálculo, la frecuencia de medición, el responsable institucional principal y una meta sugerida. Presenta el resultado como tabla, con una columna por dimensión, KPI, fórmula/unidad, frecuencia, responsable principal y meta sugerida” a partir del contenido conceptual y normativo desarrollado por la autora.

El tablero de mando anterior es una herramienta de gestión que ayuda a identificar problemas, priorizar acciones correctivas y documentar pruebas (actas, registros, auditorías e informes) que mejoran la trazabilidad del consentimiento informado y la legitimidad del uso de IA en la institución.

4. 7. KPI de Consentimiento, Confianza y Auditabilidad

Para establecer los KPI de consentimiento, se retoma el componente de confianza del MGC. En el contexto de las IES en México, el carácter se expresa, entre otras prácticas, en evitar condicionalidades indebidas y garantizar ausencia de perjuicio por negarse o retirar consentimiento y habilitar vías efectivas de remedio; mientras que la competencia se evidencia mediante trazabilidad técnica, claridad informativa, integridad del registro probatorio y formación permanente para supervisar y cuestionar usos de IA (Covey, 2006; U. S. *Department of Education*, 2023). En contraposición, la baja confianza se traduce en fricción institucional cuando existen conductas no éticas o cuando hay buenas intenciones sin capacidades suficientes para ejecutar controles consistentes (Covey, 2006)¹⁴. En consecuencia, la Tabla 11 presenta un conjunto de KPIs comparables entre unidades académicas, centrados en consentimiento y confianza, diseñados para auditorías periódicas basadas en evidencia y para la mejora continua de procesos donde interviene IA.

Asimismo, estos KPIs se fundamentan en los criterios del Comité Europeo de Protección de Datos para un consentimiento válido —libre, específico, informado e inequívoco— y en sus exigencias operativas de demostrabilidad, facilidad de retiro y adecuación del consentimiento como base jurídica sólo cuando existe elección real (Comité Europeo de Protección de Datos [CEPD], 2020). En procesos con efectos significativos (admisión, becas, permanencia, acreditación o sanciones), el consentimiento se concibe como parte de un paquete de garantías que integra, como mínimo, consentimiento dinámico (capacidad de actualizar preferencias y retirar sin fricción), evaluaciones de impacto y medidas de control, así como garantías procedimentales (revisión humana significativa, impugnación y trazabilidad de responsabilidades), en coherencia con la lógica de marcos de IA fiable y gestión iterativa de riesgos (Grupo de Expertos de Alto Nivel sobre IA, 2019; *NIST*, 2023).

En la Tabla 11 se muestra como el MGC propuesto convierte la confianza en un criterio que articula estructura institucional, control humano y controles verificables.

¹⁴ Recordemos que Covey señala que hay una distinción entre carácter (integridad, motivos e intención) y competencia (capacidades, habilidades y resultados verificable) Vid. Covey, 2006.

Tabla 11. KPI para implementar el MGC.

#	Dimensión MGC	KPI	Fórmula/unidad	Frecuencia	Meta sugerida	Covey: Confianza
1	Gobernanza	Cobertura de evaluación de "libertad" (no detrimento) en tratamientos basados en consentimiento	$(N \text{ tratamientos con checklist de libertad completo} / N \text{ total tratamientos basados en consentimiento}) \times 100$	Trimestral	$\geq 95\%$	Carácter
2	Operativa (técnico-operativa)	Granularidad del consentimiento por finalidad (desagregación)	$(N \text{ flujos con finalidades separadas} / N \text{ total flujos que requieren consentimiento}) \times 100$	Semestral	$\geq 90\%$	Competencia
3	Gobernanza	Separación entre condiciones del servicio y consentimiento para tratamientos no necesarios	$(N \text{ flujos conformes} / N \text{ total flujos evaluados}) \times 100$	Trimestral	100%	Carácter
4	Gobernanza + Pedagógica	Complejidad del paquete informativo en el momento de la solicitud	$(N \text{ flujos con paquete informativo completo} / N \text{ total flujos que requieren consentimiento}) \times 100$	Semestral	$\geq 95\%$	Ambos
5	Operativa (técnico-operativa)	Acción afirmativa inequívoca (sin casillas premarcadas ni ambigüedad)	$(N \text{ flujos con opt-in inequívoco} / N \text{ total flujos evaluados}) \times 100$	Trimestral	100%	Competencia
6	Operativa (técnico-operativa) + Pedagógica	Simetría de retiro del consentimiento (es tan fácil retirar como otorgar)	Índice = $(\text{Pasos para retirar} / \text{Pasos para otorgar})$	Trimestral	≤ 1.2	Ambos
7	Operativa (técnico-operativa)	Cumplimiento de plazo de retiro y cese del tratamiento	$(N \text{ retiros ejecutados dentro del SLA} / N \text{ retiros solicitados}) \times 100$	Mensual	$\geq 95\%$	Competencia
8	Gobernanza + Operativa (técnico-operativa)	Recambio/renovación de consentimiento ante cambios sustantivos	$(N \text{ cambios de consentimiento aplicable cumplidos} / N \text{ cambios sustantivos identificados}) \times 100$	Trimestral	$\geq 95\%$	Competencia
9	Operativa (técnico-operativa)	Trazabilidad del consentimiento (registro verificable)	$(N \text{ registros completos} / N \text{ consentimientos registrados}) \times 100$	Mensual	$\geq 99\%$	Competencia
10	Gobernanza	Cobertura de Evaluación de impacto (DPIA/Al-Impact) en procesos de alto impacto	$(N \text{ casos IA alto impacto con evaluación vigente} / N \text{ total casos IA alto impacto}) \times 100$	Semestral	100%	Competencia
11	Gobernanza + Pedagógica	Notificación efectiva del uso de IA en decisiones significativas	$(N \text{ decisiones conforme} / N \text{ total decisiones significativas con IA}) \times 100$	Mensual	$\geq 95\%$	Carácter
12	Gobernanza + Operativa (técnico-operativa)	Revisión humana significativa (meaningful human oversight) disponible y usada	Índice = $0.5 \times (\% \text{ con opción de revisión}) + 0.5 \times (\% \text{ muestreo con evidencia de revisión})$	Trimestral	$\geq 90\%$	Ambos
13	Gobernanza	Tasa de impugnación/resolución y tiempos de respuesta (garantías procedimentales)	a) $(\text{Impugnaciones/Decisiones}) \times 1000$; b) por unidad; c) $(\text{Resueltas SLA/Impugnaciones}) \times 100$	Trimestral	Definir por institución	Carácter
14	Formativa + Pedagógica	Índice de comprensión del consentimiento (quiz breve post-aviso)	Puntaje promedio por unidad académica (0-100) + % que alcanza umbral ≥ 80	Semestral	Promedio ≥ 80	Competencia
15	Formativa	Cobertura de alfabetización en IA y privacidad para personal decisor y revisor	$(N \text{ personal capacitado} / N \text{ total personal clave}) \times 100$	Anual (seguimiento trimestral)	$\geq 90\%$	Ambos
16	Gobernanza + Pedagógica + Formativa	Índice de confianza estudiantil en procesos con IA (instrumento estandarizado)	Promedio 0-100 + varianza entre unidades (brecha)	Anual (con corte semestral si hay cambios)	Definir por institución (benchmark)	Ambos
17	Gobernanza	Tasa de quejas/reclamos vinculados a consentimiento e IA (por 10,000 usuarios)	$(\text{Quejas en el periodo} / N \text{ usuarios}) \times 10,000$	Trimestral	Tendencia a la baja	Carácter
18	Gobernanza + Operativa (técnico-operativa)	Cierre de acciones correctivas derivadas de auditorías de consentimiento/IA	$(\text{Hallazgos cerrados en plazo} / \text{Hallazgos totales}) \times 100$; + % reincidencia	Trimestral	$\geq 90\%$ en plazo; reincidencia ↓	Competencia
19	Gobernanza + Operativa (técnico-operativa)	Uso correcto de consentimiento explícito cuando se tratan categorías especiales de datos	$(N \text{ flujos con consentimiento explícito conforme} / N \text{ total flujos con datos especiales basados en consentimiento}) \times 100$	Semestral	100%	Ambos

Nota. Elaboración propia con apoyo de Claude Sonnet 5 (Anthropic, 2026), a partir del prompt "A partir de este texto, genera una tabla de KPI para implementar el Modelo de Gobernanza para la Confianza (MGC), enfocada específicamente en el consentimiento informado. Organiza las columnas como: número, dimensión

del MGC a la que corresponde cada indicador (Gobernanza, Operativa, Pedagógica, Formativa, o combinaciones entre ellas), el KPI, la fórmula o unidad de medición, la frecuencia de medición, la meta sugerida, y una columna adicional que indique si el indicador corresponde al carácter o a la competencia del modelo de confianza de Covey (2006), o a ambos. Incluye indicadores que cubran: evaluación de la libertad y ausencia de detrimento en el consentimiento, granularidad por finalidad, separación entre condiciones del servicio y consentimiento, completitud del paquete informativo, acción afirmativa inequívoca, simetría entre otorgar y retirar el consentimiento, cumplimiento de plazos de retiro, trazabilidad, evaluación de impacto en procesos de alto riesgo, notificación de uso de IA en decisiones significativas, revisión humana significativa, impugnación y tiempos de respuesta, comprensión del consentimiento, alfabetización en IA del personal, confianza estudiantil, quejas y reclamos, cierre de acciones correctivas de auditorías, y uso correcto del consentimiento explícito para categorías especiales de datos" y del contenido conceptual y normativo desarrollado por la autora.

La Tabla 11 funciona como puente operativo entre la arquitectura conceptual del MGC y la implementación: traduce principios en métricas auditables, comparables y orientadas a la mejora continua, con evidencia suficiente para sostener responsabilidad proactiva (CEPD, 2020).

Con el fin de no sobrecargar el cuerpo del capítulo con especificaciones instrumentales, las fichas técnicas de los KPIs se presentan en el Anexo 1, organizadas en dos tablas: (i) Tabla A, correspondiente al tablero KPI del MGC (relacionada con la Tabla 10), y (ii) Tabla B, correspondiente a los KPIs de consentimiento, confianza y auditabilidad (relacionada con la Tabla 11). Este anexo funciona como soporte metodológico y operativo para la adopción del modelo, al establecer criterios uniformes de medición y responsabilidades institucionales, sin alterar la lógica argumentativa del capítulo.

Por todo lo antes presentado en esta investigación, se considera que la propuesta de consentimiento informado —dentro del MGC— fortalece y supera el consentimiento meramente formal. En la era de la IA en la educación superior, la privacidad del estudiantado no se protege únicamente con un documento al inicio de un curso o con una casilla de verificación digital. El consentimiento auténtico debe operar como un proceso educativo, dialógico y continuo que debe operar en las dimensiones de Gobernanza, Operativa, Pedagógica y Formativa. Solo una perspectiva holística y humanista de la gestión del consentimiento informado generará confianza en la IES mexicanas.

Conclusiones

Los hallazgos de esta investigación permiten sostener que el uso intensivo de sistemas de IA —particularmente generativos y agénticos— en las IES tensiona el equilibrio entre la innovación educativa y el derecho a la privacidad del estudiantado, y que esa tensión no se resuelve mediante el consentimiento tal como hoy se recaba. Se confirma así la brecha entre licitud y protección efectiva: el consentimiento informado -instrumentado en el aviso de privacidad- es insuficiente para expresar una voluntad genuina. Las conclusiones que siguen responden ordenadamente a las cuatro preguntas de investigación, exponen las aportaciones alcanzadas y precisan los límites del estudio.

La IA se concreta en la educación superior mediante aplicaciones y sistemas cuyo funcionamiento depende íntegramente de la disponibilidad y circulación de datos. El análisis realizado evidencia que su incorporación acelerada vuelve imprescindible delimitar qué se recolecta, con qué finalidad y en qué condiciones de legitimidad, transparencia y rendición de cuentas, lo que sitúa el problema como un asunto de gobernanza universitaria de datos y algoritmos antes que como una cuestión meramente tecnológica.

La investigación mostró además que el tránsito de la IA generativa a la agéntica modifica sustancialmente el alcance de esa gobernanza. Mientras los sistemas generativos plantean principalmente desafíos relativos a la integridad académica, la calidad, la veracidad y los sesgos de los contenidos producidos, los sistemas agénticos incorporan riesgos derivados de su capacidad para planificar, ejecutar acciones y decidir con grados variables de autonomía. En consecuencia, la protección de los derechos de la comunidad universitaria exige mecanismos de supervisión humana significativa, trazabilidad, rendición de cuentas y consentimiento dinámico que trascienden los modelos tradicionales de protección de datos, concebidos para tratamientos estáticos y finalidades predeterminadas.

La revisión de los marcos éticos permitió identificar diez principios -algunos dobles-recurrentes que operan como referentes para la elaboración de un marco institucional de uso ético de la IA: centralidad de la persona, privacidad y seguridad de los datos, transparencia y explicabilidad, responsabilidad y rendición de cuentas, gobernanza y supervisión, inclusión y equidad, sostenibilidad y proporcionalidad, no maleficencia y prevención de daños, supervisión humana y alfabetización. A ellos se suman, como complementos operativos, el enfoque ACE —que sitúa al docente en el centro— (U. S. Department of Education, Office of Educational Technology, 2023) y el marco para una IA fiable (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019).

La aportación propia del capítulo segundo consistió en dotar a esos principios de un fundamento antropológico explícito mediante el Triángulo de la Irreductibilidad Humana. La empatía, el juicio ético y el pensamiento divergente constituyen capacidades que los sistemas de IA no pueden sustituir y que, por ello, delimitan el perímetro de lo que puede delegarse

en ellos. Su relevancia excede el plano conceptual: el Triángulo opera como criterio de gobernanza institucional, en la medida en que de él se sigue que toda decisión con impacto significativo sobre estudiantes o docentes debe preservar la deliberación y la responsabilidad humanas. La segunda conclusión de este bloque es que la confianza constituye el eje transversal capaz de articular principios éticos y mecanismos operativos, y que sólo adquiere valor jurídico cuando se traduce en evidencias verificables.

El contraste entre los tres ordenamientos jurídicos que se realizó en el capítulo tres, mediante una matriz de dieciocho criterios agrupados en seis ejes de análisis —arquitectura normativa, posición del consentimiento en la estructura de licitud, estándar de validez, tratamiento específico de la IA, cadena de proveedores y arquitectura de garantía—, permitió formular tres hallazgos. El primero es de convergencia conceptual: la UE, EUA y México definen el consentimiento en términos casi idénticos, de modo que la divergencia no se sitúa en el concepto sino en la carga que cada uno hace recaer sobre él. El segundo es de divergencia funcional: la UE ha desplazado expresamente la protección desde la voluntad individual hacia las obligaciones de diseño, las prohibiciones absolutas y la gobernanza del despliegue —de ahí que prohíba la inferencia de emociones en centros educativos al margen de toda autorización individual—; EUA opera ese desplazamiento de manera parcial, por vía contractual y de mercado, con un régimen educativo federal anterior a internet y desprovisto de acción privada; y México conserva un modelo centrado en el consentimiento y el aviso de privacidad sin haber articulado los instrumentos llamados a sustituirlo. El tercero es de inversión entre exigencia y garantía: allí donde la exigencia sustantiva mexicana alcanza su punto más alto —el régimen de los sujetos obligados—, la garantía institucional alcanza el más bajo, pues la función garante se ha internalizado en el órgano interno de control de la propia institución supervisada.

El ordenamiento mexicano tiene comprende omisiones que queremos destacar: la ausencia de un régimen jurídico de la IA, la debilidad del derecho de oposición frente al tratamiento automatizado, que es reactivo y carece de derechos correlativos a la intervención humana, a la explicación y a la impugnación, y un estándar de validez del consentimiento inferior al comparado, que omite el requisito de que la manifestación sea inequívoca y admite la modalidad tácita como regla general. A ellas se suma la ausencia de una norma sectorial educativa útil, pues la Ley General de Educación resulta prácticamente muda en materia de consentimiento estudiantil e inaplicable al alumnado mayor de edad.

El análisis de los regímenes de consentimiento no estaría completo sin la dimensión institucional desarrollada en el apartado 3.6.6: la extinción del Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales no fue seguida de un órgano equivalente, sino de una redistribución de competencias que internalizó la función garante —en la Secretaría Anticorrupción y Buen Gobierno para los particulares, y en el órgano interno de control de cada sujeto obligado para las instituciones públicas—. De ahí se sigue la conclusión más relevante del análisis del caso mexicano: la inversión entre

exigencia sustantiva y garantía institucional. El régimen aplicable a las instituciones públicas es materialmente más estricto que el aplicable a las privadas y, sin embargo, queda confiado a una supervisión que se ha vuelto interna, mientras que el régimen privado, más permisivo en sustancia, conserva al menos formalmente una autoridad externa. El contraste con el ordenamiento europeo —donde fueron precisamente las autoridades independientes quienes advirtieron sobre los riesgos del paquete Digital Omnibus (EDPB & EDPS, 2026)— confirma que la existencia de un órgano externo es condición de posibilidad de la crítica institucional, y que la divergencia entre ambos ordenamientos no radica en el contenido de los derechos reconocidos, sino en la arquitectura destinada a hacerlos exigibles.

De este diagnóstico se sigue la conclusión central del capítulo tercero: en el estado actual del ordenamiento mexicano, los marcos institucionales de uso ético de la IA no operan como complemento voluntario de la normativa, sino como la principal garantía efectivamente disponible. Ello se sostiene en tres razones de necesidad —el desfase entre el ciclo legislativo y el tecnológico, la imposibilidad estructural de que el consentimiento soporte la carga que se le atribuye y el hecho de que, donde la ley calla, la norma institucional es la única fuente de obligación— y en cuatro razones de idoneidad: el momento jurídicamente decisivo es la contratación del proveedor y no la firma del estudiante; la internalización de la función garante en los órganos internos de control exige un parámetro interno verificable con el cual contrastar la conducta institucional, de modo que aquello que constituye una debilidad desde el punto de vista de la independencia incrementa paradójicamente la relevancia jurídica de la norma interna; la operación simultánea en varios ordenamientos aconseja un sistema de gestión que funcione como paraguas de cumplimiento neutral; y la relación educativa, sostenida en la confianza, permite a la institución ocuparse de aquello que ninguna ley de protección de datos alcanza a regular.

Aunado a lo anterior, la ENIAG 2025 aporta evidencia empírica directa sobre la brecha entre la existencia formal de normativa -el consentimiento- y su conocimiento efectivo por parte de la comunidad universitaria: más del 76% de los estudiantes y el 75% de los docentes de instituciones públicas declaran no conocer ninguna normativa institucional sobre el uso de la IA (SEP, 2025). La misma brecha, aunque algo menor, se observa en instituciones particulares, donde el desconocimiento se ubica entre el 61% y el 67% según el grupo consultado. Por lo tanto, la existencia de un aviso de privacidad o de un documento normativo no garantiza que el consentimiento informado opere realmente como tal, pues el estudiantado y el profesorado desconocen, en su mayoría, las reglas bajo las cuales sus datos son tratados.

Ante el vacío de la legislación mexicana en materia de IA, se proponen normas ISO/IEC: la ISO/IEC 27701 para la gestión de la privacidad, la ISO 27018 para la protección de datos en servicios en la nube, la ISO 29184 para los avisos de privacidad y el consentimiento, la TS 27560 para su registro verificable y la ISO 42001 para la gobernanza de sistemas de IA. Estos estándares internacionales no sustituyen el marco normativo vigente, sino que contribuyen a

hacerlo operativo, verificable y auditable mediante sistemas de gestión, controles documentados y mejora continua (ISO, 2019a, 2019b, 2020, 2023; Lachaud, 2020).

El análisis de los instrumentos institucionales expedidos por siete instituciones mexicanas entre 2023 y 2026 constató un avance significativo en la construcción de marcos de gobernanza y una convergencia apreciable en principios como la transparencia, la protección de datos, la supervisión humana y la no discriminación. Constató igualmente que la mayoría de estas iniciativas privilegia la formulación de principios y lineamientos generales, sin desarrollar mecanismos que articulen de manera sistemática las dimensiones jurídica, organizacional, operativa, pedagógica y formativa, ni instrumentos verificables para la gestión del consentimiento informado. Este hallazgo confirma el vacío enunciado en la introducción y justifica la propuesta que constituye la aportación principal de la tesis.

El Modelo de Gobernanza para la Confianza (MGC) articula el enfoque ecológico de Chan (2023), los marcos de IA fiable y gestión de riesgos (Grupo Independiente de Expertos de Alto Nivel sobre IA, 2019; NIST, 2023), la perspectiva pedagógica ACE (U. S. Department of Education, Office of Educational Technology, 2023) y la teoría de la confianza basada en carácter y competencia (Covey, 2006). Se implementa mediante cuatro dimensiones — gobernanza, operativa-tecnológica, pedagógica y formativa— cuya integración persigue fortalecer simultáneamente la estructura institucional, el dominio humano y los controles verificables. La dimensión de gobernanza establece protocolos, designa responsables y dicta políticas para garantizar un consentimiento libre e informado y proteger los derechos fundamentales del estudiante (NIST, 2023; International Organization for Standardization, 2023); la operativa incorpora la privacidad por defecto, las interfaces de transparencia y el consentimiento reforzado en capas y dinámico; la pedagógica introduce la transparencia didáctica y las pausas pedagógicas; y la formativa despliega el plan transversal de alfabetización.

Una aportación conceptual particularmente relevante consiste en reconocer la alfabetización en IA como dimensión autónoma del modelo y no como componente de la operación tecnológica. Aunque diversos marcos internacionales incluyen acciones de capacitación, la misión propia de las IES exige que la formación ética, crítica y digital constituya una función de gobernanza con objetivos, responsables e indicadores diferenciados. La dimensión formativa fortalece la autonomía de docentes y estudiantes para comprender, supervisar y cuestionar el funcionamiento de estos sistemas, y con ello hace posibles tanto la supervisión humana significativa como un consentimiento realmente informado (Dignum, 2019; UNESCO, 2021). En este sentido, la dimensión pedagógica y la formativa dejan de ser accesorias y se vuelven componentes sustantivos del consentimiento, porque son las que posibilitan la comprensión real y la deliberación.

Los aportes operativos del modelo se expresan en instrumentos y controles concretos: registro institucional de sistemas y clasificación de casos de uso por nivel de riesgo, doble

evaluación de impacto ex ante y revisión continua, transparencia reforzada, revisión humana de decisiones con efectos significativos y consentimiento dinámico. Todo ello se traduce en una matriz de operacionalización y en un cuadro de mando de indicadores (Tabla 10 y 11) que convierten la confianza en evidencia de responsabilidad proactiva, con fichas técnicas (Anexo 1) que precisan definición operativa, fórmula, fuente, periodicidad, responsable, meta y riesgo asociado.

Cabe concluir, finalmente, que la principal fortaleza del modelo MGC radica en su carácter adaptable frente a la evolución tecnológica. Al fundamentarse en principios de gobernanza, supervisión humana, gestión de riesgos y mejora continua, resulta aplicable tanto a los desafíos actuales de la IA generativa como a los escenarios emergentes asociados con la agéntica, cuya creciente autonomía exige mecanismos más robustos de trazabilidad, consentimiento dinámico y rendición de cuentas.

Los hallazgos sostienen la proposición de trabajo enunciada en la introducción: cuando el consentimiento informado se reduce a avisos formales, sin mecanismos operativos de comprensión, trazabilidad y control en contextos de IA, la tutela de la privacidad se debilita. El análisis comparado mostró que el ordenamiento europeo ha extraído esa misma conclusión y ha reaccionado desplazando la protección hacia el diseño, la prohibición y la gobernanza; que el estadounidense la ha desplazado hacia el contrato y el mercado; y que el mexicano ha conservado la retórica del consentimiento habiendo renunciado a sus condiciones de eficacia.

Queda con ello acreditada la tesis defendida: en una relación estructuralmente asimétrica, y ante un ordenamiento que no ha articulado sustitutos, la garantía efectiva debe trasladarse desde el acto de voluntad individual hacia la gobernanza institucional verificable. La contribución de esta investigación consiste en demostrar que la confianza puede concebirse como criterio institucional verificable y no únicamente como valor abstracto, y en ofrecer el instrumento que permite hacerlo: el MGC se presenta como un modelo que articula dimensiones jurídicas, organizacionales, operativas, pedagógicas y formativas, y que se propone como estándar de debida diligencia institucional.

Del conjunto de la investigación se derivan recomendaciones diferenciadas según su destinatario.

A las IES: adoptar marcos vinculantes y no meramente declarativos, incorporados a la normativa institucional y articulados con las políticas de contratación y compras; constituir un órgano colegiado responsable del uso ético de la IA; levantar un registro de sistemas con clasificación por nivel de riesgo; y establecer cláusulas contractuales tipo que delimiten expresamente el uso de los insumos institucionales para el entrenamiento o la mejora de los modelos del proveedor.

A los órganos internos de control y a las instancias de auditoría: asumir que la función garante internalizada sólo puede ejercerse sobre un parámetro previamente establecido, por lo que la

existencia de indicadores verificables de consentimiento y trazabilidad constituye condición de posibilidad de su propia actuación.

A las asociaciones y organismos del sistema de educación superior: promover criterios comunes y comparables entre instituciones, de modo que la gobernanza de la IA no dependa exclusivamente de la capacidad instalada de cada universidad y no profundice las asimetrías existentes entre ellas.

Al legislador mexicano: considerar la incorporación de un régimen específico para los tratamientos automatizados de alto impacto en el ámbito educativo, el restablecimiento del deber de informar los destinatarios de las transferencias, la regulación legal de la relación con encargados y proveedores de servicios en la nube en el régimen de los particulares, y el reconocimiento de la asimetría de poder como criterio de valoración de la libertad del consentimiento.

La investigación presenta limitaciones que conviene explicitar. En primer lugar, el MGC constituye una propuesta fundamentada teórica y normativamente, pero no ha sido sometido a validación empírica ni a implementación piloto, por lo que su viabilidad operativa y sus efectos sobre la comprensión efectiva del estudiantado permanecen como hipótesis por contrastar. En segundo lugar, la muestra de instituciones analizadas es intencional y no representativa: se seleccionó por diversidad de régimen y disponibilidad pública del instrumento, de modo que sus resultados describen tendencias y no permiten inferencias sobre el conjunto del sistema de educación superior mexicano; asimismo, las categorías recogen únicamente formulaciones documentadas en las fuentes consultadas, sin suponer exhaustividad respecto del contenido íntegro de cada instrumento. Finalmente, el diseño es documental y no incorpora datos primarios sobre la percepción, la comprensión o la experiencia del estudiantado y del profesorado, perspectiva que resultaría necesaria para valorar la eficacia real de los mecanismos propuestos.

De las limitaciones anteriores se derivan tres líneas de continuidad. La primera consiste en el pilotaje del modelo en una o varias instituciones, con medición longitudinal de los indicadores propuestos, que permita evaluar su viabilidad, su costo organizacional y su efecto sobre la calidad del consentimiento. La segunda, de carácter empírico, apunta al estudio de la comprensión efectiva del estudiantado respecto de los avisos y mecanismos de consentimiento en entornos mediados por IA, con metodologías mixtas que permitan contrastar comprensión declarada y comprensión real. La tercera propone ampliar y actualizar el análisis de instrumentos institucionales, extendiéndose a un número mayor de instituciones y a otros sistemas de educación superior de América Latina, a fin de identificar patrones regionales de gobernanza.

Gobernar la IA en la universidad implica, al mismo tiempo, proteger los datos, integrar los sistemas y desarrollar las capacidades humanas. Solo un modelo que articule las dimensiones jurídica, organizacional, operativa, pedagógica y formativa permitirá que estas tecnologías contribuyan al desarrollo de una educación superior ética, responsable y orientada al bien común.

La universidad que gobierna datos y algoritmos con transparencia, y que forma a su comunidad para comprender sus efectos, convierte la innovación educativa y la incorporación de la IA en un proceso legítimo, responsable y compatible con la dignidad de la persona y la promoción del bien común.

Bibliografía

- About Ali, M., Dornaika, F., & Charafeddine, J. (2025). Agentic AI: A comprehensive survey of architectures, applications, and future directions. *Artificial Intelligence Review*, 59(1), Artículo 11. <https://doi.org/10.1007/s10462-025-11422-4>
- Ahmad, N., Murugesan, S., & Kshetri, N. (2023). Generative artificial intelligence and the education sector. *Computer*, 56 (6), 72–86. <https://doi.org/10.1109/MC.2023.3263576>
- Alonso-Rodríguez, A. M. (2024). Hacia un marco ético de la inteligencia artificial en la educación. *Teoría de la Educación. Revista Interuniversitaria*, 36(2), 79-98. <https://doi.org/10.14201/teri.31821>
- Alpaydin, E. (2016). *Machine Learning: The New AI*. Cambridge, MA: MIT Press.
- Álvarez Caro, M. (2015). Derecho al olvido en internet: El nuevo paradigma de la privacidad en la era digital. Editorial Reus. <https://www.editorialreus.es/libros/derecho-al-olvido-en-internet-el-nuevo-paradigma-de-la-privacidad-en-la-era-digital/9788429018363/>
- Angeli, Ch. et al. (2015). Processing Natural Language in AI Systems. *IEEE Transactions on Learning Technologies*.
- Arendt, H. (1993). *La condición humana*. Paidós.
- Aristóteles. (2000). *Ética a Nicómaco*. Gredos.
- Arnold, K. & Pistilli, M. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270. <https://doi.org/10.1145/2330601.2330666>
- Asociación Nacional de Universidades e Instituciones de Educación Superior [ANUIES]. (2025). *Gobernanza de la inteligencia artificial...* (J. L. Ponce López y C. H. de J. Díaz Novelo, Coords.). ANUIES.

- Asociación por los Derechos Civiles (2016). El sistema de protección de datos personales en América Latina: oportunidades y desafíos para los derechos humanos (Vol. I). <https://adc.org.ar/informes/sistema-proteccion-datos-personales-latam/>
- Baker, R & Siemens, G. (2014). Educational data mining and learning analytics. *Cambridge Handbook of the Learning Sciences*, 2, 253-272. <https://doi.org/10.1017/CBO9781139519526.016>
- Baker, R. & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3-17. <https://jedm.educationaldatamining.org/index.php/JEDM/article/view/8>
- Ballart, X. (1993). Evaluación de políticas. Marco conceptual y organización institucional. *Revista de Estudios Políticos (Nueva Época)*, (80), 199–224. <https://www.cepc.gob.es/publicaciones/revistas/revista-de-estudios-politicos/numero-80-abriljunio-1993/evaluacion-de-politicas-marco-conceptual-y-organizacion-institucional-1>
- Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic AI: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. *Future Internet*, 17(9), Artículo 404. <https://doi.org/10.3390/fi17090404>
- Barajas-Ochoa, A. y Barajas-Ochoa, Z. (2019). Respecto al consentimiento informado y su documentación. *Gaceta Médica de México*, 155(2), 217. <https://doi.org/10.24875/GMM.19005031>
- Barrio Andrés, M. (2022). La regulación del derecho a la protección de datos en los Estados Unidos: hacia un RGPD norteamericano. *Cuadernos de Derecho Transnacional*, 14(2), 186–193. <https://doi.org/10.20318/cdt.2022.7181>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., & Liang, P. (2021). *On the opportunities and risks of foundation models* (arXiv:2108.07258). arXiv. <https://doi.org/10.48550/arXiv.2108.07258>

- Borgman, C. L. (2018). Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, 33(2), 365–412. <https://doi.org/10.15779/Z38B56D489>
- Brown et al. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. -F. Balcan, & H. -T. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html>
- Burger, B., Maffettone, P. M., Gusev, V. V., Aitchison, C. M., Bai, Y., Wang, X., Li, X., Alston, B. M., Li, B., Clowes, R., Rankin, N., Harris, B., Sprick, R. S., & Cooper, A. I. (2020). A mobile robotic chemist. *Nature*, 583(7815), 237–241. <https://doi.org/10.1038/s41586-020-2442-2>
- Bustelo Gracia, J. L. (2025). Hacia una gobernanza algorítmica transparente: auditoría de sesgo. Estudio de caso. *Cuadernos de Gobierno y Administración Pública*, 12(1), e97604. <https://doi.org/10.5209/cgap.97604>
- Canty, R. B., Bennett, J. A., Brown, K. A., Buonassisi, T., Kalinin, S. V., Kitchin, J. R., & Abolhasani, M. (2025). Science acceleration and accessibility with self-driving labs. *Nature Communications*, 16, Article 3856. <https://doi.org/10.1038/s41467-025-59231-1>
- Castañeda, L. (2013). *Entornos personales de aprendizaje: Claves para el ecosistema educativo en red*. Marfil EBooks.
- Castells, M. (2006). *La sociedad red: una visión global*. Alianza Editorial.
- Cavoukian, A. (2012). *Operationalizing privacy by design: A guide to implementing strong privacy practices*. Information and Privacy Commissioner of Ontario. <https://gpsbydesigncentre.com/wp-content/uploads/2021/08/Doc-5-Operationalizing-pbd-guide.pdf>
- Cinquegrana, P., & Aggarwal, S. (2018). Deep learning: The latest trend in AI and ML. Qubole. <https://www.qubole.com/blog/deep-learning-the-latest-trend-in-ai-and-ml>

- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20, Article 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Comandé G., Editor (2022). *Elgar encyclopedia of law and data science*. Edward Elgar Publishing. <https://www.e-elgar.com/shop/gbp/elgar-encyclopedia-of-law-and-data-science-9781839104589.html>
- Cordón García, O. (2023). Inteligencia Artificial en Educación Superior: Oportunidades y Riesgos. *RiiTE Revista interuniversitaria de investigación en tecnología educativa*, 15, 16-27. <https://doi.org/10.6018/riite.591581>
- Cornejo-Plaza, I. y Cippitani, R. (2023). Consideraciones éticas y jurídicas de la IA en educación superior. *Revista de Educación y Derecho*, 28(1), 1-10. <https://doi.org/10.1344/REYD2023.28.43935>
- Covey, S. M. R., & Merrill, R. R. (2006). *The speed of trust: The one thing that changes everything*. Free Press.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(22). <https://link.springer.com/article/10.1186/s41239-023-00392-8>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Ding, L., Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies*, 29, 14151–14203. <https://doi.org/10.1007/s10639-023-12402-3>

- Doshi-Vélez, F., & Kim, B., (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv: 1702. 08608. <https://arxiv.org/abs/1702.08608>
- Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 211–407. <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>
- Ed. Schoeman (1984). *Philosophical Dimensions of Privacy: An Anthology*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511625138>
- Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a "right to an explanation" is probably not the remedy you are looking for. *Duke Law & Technology Review*, 16(1), 18–84. <https://doi.org/10.31228/osf.io/97upg>
- Elsevier. (2025). *A guide to agentic AI for academic institutions*. <https://www.elsevier.com/academic-and-government/ai-in-research-and-higher-education/how-to-harness-agentic-ai-in-academia>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- European Digital Rights. (2026). *The Digital Omnibus: A step back from the brink, but the risks remain*. <https://edri.org/>
- Faiad, S. L. (2022). La ética en los ensayos clínicos: de Tuskegee al Informe Belmont. *Oncología Clínica*, 27(2). <https://doi.org/10.56969/oc.v27i2.83>
- Farber, S. (2025). Enhancing academic decision-making: A pilot study of AI-supported journal selection in higher education. *Innovative Higher Education*. <https://doi.org/10.1007/s10755-025-09791-3>
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5–6), 304–317. <https://doi.org/10.1504/IJTEL.2012.051816>
- Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, 66(1), 111–126. <https://doi.org/10.1007/s12599-023-00834-7>

- Flores-Vivar, J.-M., y García-Peñalvo, F. J. (2023). Reflexiones sobre la ética, potencialidades y retos de la inteligencia artificial en el marco de la educación de calidad (ODS4). *Comunicar*, 31(74), 37–47. <https://doi.org/10.3916/C74-2023-03>
- Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information Technology*, 8(3), 109–119. <https://doi.org/10.1007/s10676-006-9121-3>
- Floridi, L. (2013). *The ethics of information*. Oxford University Press.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Frankl, V. (2004). *El hombre en busca de sentido*. Herder. (Obra original publicada en 1946).
- Freeman, CH., & Louçã, F. (2002). *As Time Goes By: From the Industrial Revolutions to the Information Revolution*. Oxford University Press. <https://doi.org/10.1093/0199251053.001.0001>, accessed 11 Aug. 2026.
- Fried, C. (1968). Privacy. *The Yale Law Journal*, 77(3), 475. <https://doi.org/10.2307/794941>
- Froomkin, A. M. (2000). The Death of Privacy? *Stanford Law Review*, 52(5), 1461. <https://doi.org/10.2307/1229519>
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2015). Learning Analytics Should Not Promote One Size Fits All: The Effects of Instructional Conditions in Predicting Academic Success. *The Internet and Higher Education*, 28, 68-84. <https://doi.org/10.1016/j.iheduc.2015.10.002>
- Gall, M., Gall, J., & Borg, R. (2007). *Educational Research: An Introduction* (8th ed.). Pearson.

- Garfinkel, S. & Bowen, C. (2022). Preserving privacy while sharing data. MIT Sloan Management Review, 63(3), 1–4.
<https://shop.sloanreview.mit.edu/preserving-privacy-with-shared-data>
- Georgieva, M., Webb, J., Stuart, J., Bell, J., Crawford, S., & Ritter-Guth, B. (2025, 24 de junio). AI ethical guidelines. EDUCAUSE.
<https://library.educause.edu/resources/2025/6/ai-ethical-guidelines>
- González, L. D. (2018). Control de nuestros datos personales en la era del big data: El caso del rastreo web de terceros. In Estudios Socio-Jurídicos (Vol. 21, Issue 1). <http://www.scielo.org.co/pdf/esju/v21n1/0124-0579-esju-21-01-209.pdf>
- González-Campos, J., López-Núñez, J., y Araya-Pérez, C. (2024). Educación superior e inteligencia artificial: Desafíos para la universidad del siglo XXI. *Aloma: Revista de Psicología, Ciències de l'Educació i de l'Esport*, 42(1), 79–90. <https://doi.org/10.51698/aloma.2024.42.1.79-90>
- González Sánchez, J. L., Villota Garcia, F. R., Moscoso Parra, A. E., Garces Calva, S. W., y Bazurto Arévalo, B. M. (2023). Aplicación de la inteligencia artificial en la educación superior. *Dominios de las Ciencias*, 9(3), 1097–1108. 10.23857
- Goodfellow, I., Bengio, Y., Courville, A., & Bach, F. (2016). Deep Learning. MIT Press.
- Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, 15 (3), 42–57. (PDF) *Translating Learning into Numbers: A Generic Framework for Learning Analytics*. Available from: https://www.researchgate.net/publication/234057371_Translating_Learning_into_Numbers_A_Generic_Framework_for_Learning_Analytics [accessed Aug 23 2026].
- Gunning D., & Aha, D. W., (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44–58.
<https://doi.org/10.1609/aimag.v40i2.2850>

- Hintze, A. (2016, November 14). *Understanding the four types of AI, from reactive robots to self-aware beings*. The Conversation. <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Holtz, H. M., & Ledendal, J. (2026). AI data governance – Overlaps between the AI Act and the GDPR. *Law, Innovation and Technology*, 380–409. <https://doi.org/10.1080/17579961.2026.2633677>
- Jones, K. M. L. (2019). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of Educational Technology in Higher Education*, 16(1). 10.1186/s41239-019-0155-0
- Jones KML, Rubel A, & LeClere E. A., matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics. *Journal of the Association for Information Science and Technology*. 2020; 71: 1227–1241. <https://doi.org/10.1002/asi.24327>
- Jurafsky, D., & Martin, J. H., (2020). *Speech and Language Processing* (3rd ed.). Pearson.
- Kairouz P, & McMahan, H.B., (2021). Advances and Open Problems in Federated Learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210. <https://doi.org/10.1561/22000000083>
- Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability*, 15(16), 12451. <https://doi.org/10.3390/su151612451>

- Kanobel, M. C. (2026). Hacia una gobernanza responsable de la inteligencia artificial generativa en la educación superior. *Revista Educación Superior y Sociedad (ESS)*, 37(2), 342–371. <https://doi.org/10.54674/ess.v37i2.1204>
- Kant, I. (2003). *Fundamentación de la metafísica de las costumbres*. Alianza. (Obra original publicada en 1785).
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics*, 23(2), 141–146. <https://doi.org/10.1038/ejhg.2014.71>
- Kingma, D. P., y Welling, M. (2022). *Auto-encoding variational Bayes* (arXiv:1312.6114v11). arXiv. <https://doi.org/10.48550/arXiv.1312.6114>
- Kisaecanin., B Bhattacharyya, S. S. & Chai, S. (2009). *Embedded computer vision*. Advances in Pattern Recognition Series. Springer-Verlag London Limited. <https://doi.org/10.1007/978-1-84800-304-0>
- Kotsiantis Pierrakeas, C., & Pintelas, P. (2013). Predicting student performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5), 411–426. <https://doi.org/10.1080/08839510490442058>
- Krumm, A. E., Means, B., & Bienkowski, M. (2018). *Learning analytics goes to school: A collaborative approach to improving education*. Routledge. <https://doi.org/10.4324/9781315650722>
- Kwet, M. y Prinsloo, P. (2020). The ‘smart’ classroom: A new frontier in the age of the smart university. *Teaching in Higher Education*, 25(4), 510–526. <https://doi.org/10.1080/13562517.2020.1734922>

- Lachaud, E. (2017). The General Data Protection Regulation and the rise of certification as a regulatory instrument. *International Review of Law, Computers & Technology*, 32(1), 26–44. <https://doi.org/10.1080/13600869.2017.1298502>
- Lachaud, E. (2020). ISO/IEC 27701: Threats and opportunities for GDPR certification. *European Data Protection Law Review*, 6(2), 194–210. <https://doi.org/10.21552/edpl/2020/2/7>
- Lanier, J. (2018). *Diez razones para borrar tus redes sociales de inmediato* (M. Vaquero, Trad.). Debate. (Obra original publicada en 2018).
- Li, X., Grandvalet, Y., Davoine, F., Cheng, J., Cui, Y., Zhang, H., Belongie, S., Tsai, Y.-H., & Yang, M.-H. (2020). Transfer learning in computer vision tasks: Remember where you come from. *Image and Vision Computing*, 93. <https://doi.org/10.1016/j.imavis.2019.103853>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. En *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Marín, V. I., & Tur, G. (2023). La privacidad de los datos en tecnología educativa: Resultados de una revisión de alcance. *EduTec. Revista Electrónica de Tecnología Educativa*, 83, 7–29. <https://doi.org/10.21556/edutec.2023.83.2701>
- Martínez Molina, O. A. (2024). Explicabilidad y educación superior: Retos éticos en la inteligencia artificial. *Revista Iberoamericana de Educación Digital*, 15(1), 33–51.
- Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: La revolución de los datos masivos* (A. J. Iriarte Jurado, Trad.). Madrid: Turner.
- Mendoza Enríquez, O.A., (2018). Marco jurídico de la protección de datos personales en las empresas de servicios establecidas en México: desafíos y cumplimiento. *REVISTA IUS*, 12(41), 267–291. <https://doi.org/10.35487/rius.v12i41.2018.355>

- Mendoza Enríquez, O. A. (2021). El derecho de protección de datos personales en los sistemas de IA. *Revista Del Instituto De Ciencias Jurídicas De Puebla*, 14, 179–207. <https://www.scielo.org.mx/pdf/rius/v15n48/1870-2147-rius-15-48-179.pdf>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv: 1301. 3781*. <https://doi.org/10.48550/arXiv.1301.3781>
- Min-Allah, N., & Alrashed, S., (2020). Smart campus - A sketch. *Sustainable Cities and Society*, 59. <https://doi.org/10.1016/j.scs.2020.102231>
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill. <https://www.cs.cmu.edu/~tom/mlbook.html>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society*, 3(2), 205395171667967. <https://doi.org/10.1177/2053951716679679>
- Mokyr, J. (1990). *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press.
- Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S. R., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, 29(1), 142–163. <https://doi.org/10.1080/10494820.2018.1558257>
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>

- Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32–48.
<https://nissenbaum.tech.cornell.edu/papers/Contextual%20Approach%20to%20Privacy%20Online.pdf>
- NOYB. (2026). Digital Omnibus report v3: Analysis of select GDPR and ePrivacy proposals by the Commission. <https://noyb.eu/>
- O'Neil, C. (2017). *Armas de destrucción matemática: Cómo el big data aumenta la desigualdad y amenaza la democracia* (V. Campos, Trad.). Capitán Swing. (Obra original publicada en 2016).
- Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1), 128–147.
<https://doi.org/10.1080/1369118X.2018.1486870>
- Osorio Sanabria, M. A., Millán Otero, K. L., González González, J. A., & Arango Serna, M. D. (2017). *Gestión de datos de investigación: Introducción al contexto de la e-Ciencia*. Editorial Universidad de Antioquia.
- Parker, C., & Nielsen, V. L. (2009). Corporate compliance systems: Could they make any difference? *Administration & Society*, 41(1), 3–37.
<https://doi.org/10.1177/0095399708328869>
- Parlamento Europeo y Consejo de la Unión Europea. (2016). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 (Reglamento General de Protección de Datos). <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- Pasetti, M. y Oliveira, N. de. (2024). Repensar la seguridad de la inteligencia artificial basada en la confianza: protección contra la discriminación algorítmica. *Veritas (Porto Alegre)*, 69 (1), e45911.
<https://doi.org/10.15448/1984-6746.2024.1.45911>
- Pedreño Muñoz, A., González Gosálbez, R., Mora Illán, T., Pérez Fernández, E. del M., Ruiz Sierra, J., & Torres Penalva, A. (2024). *La inteligencia artificial en las universidades: retos y oportunidades*. Grupo 1MillionBot

- Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, 4(1). <https://doi.org/10.1093/CYBSEC/TYY001>
- Pincay Ponce, J. I., Zambrano Zambrano, M. T. ., & Quijije Anchundia, P. J. Hacia una educación personalizada y equitativa: análisis de la convergencia entre tecnologías asistidas e inteligencia artificial. *Revista Científica Multidisciplinaria SAPIENTIAE*. ISSN: 2600-6030, 8(16), 536-545.
- Porteles, L., & Jhoel, Y. (2022). Las EdTech transformando la educación universitaria. *Revista Arbitrada del CIEG*, 55, 206-215.
- Porcelli, A. M. (2019). La Protección de los Datos Personales en el Entorno Digital. Los Estándares de Protección de Datos en los Países Iberoamericanos. *Revista Quaestio Iuris*, 12(2), 465–497. <https://doi.org/10.12957/RQI.2019.40175>
- Porcelli, A. M., & Martínez, A. N. (2020). La reformulación del derecho a la privacidad y el reconocimiento de los nuevos derechos en el entorno digital en tiempos de covid-19. *Red Sociales, Revista Del Departamento de Ciencias Sociales*, 07, 109–125.
- Prinsloo, P. (2020). Data frontiers and frontiers of power in (higher) education. *Teaching in Higher Education*, 25(4), 510-526. <https://doi.org/10.1080/13562517.2020.1723537>
- Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. *ACM International Conference Proceeding Series*, 16-20-March-2015, 83–92. <https://doi.org/10.1145/2723576.2723585>
- Prinsloo, P., & Slade, S. (2018). En J. Lester, C. Klein, A. Johri, & H. Rangwala (Eds.), *Learning analytics in higher education* (pp. 118–139). Routledge.
- Quijano Decanini, C. (2022). Derecho a la privacidad en Internet. Editorial Tirant lo Blanch. <https://editorial.tirant.com/mex/autorList/carmen-quijano-decanini-579805>

- Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review*, 55(3), 2495–2527. <https://doi.org/10.1007/s10462-021-10068-2>
- Reidenberg, J. R., y Schaub, F. (2018). Achieving big data privacy in education. *Theory and Research in Education*, 16(3), 263–279. <https://doi.org/10.1177/1477878518805308>
- Reynolds, P. D. (1979). *Ethical dilemmas and social science research: An analysis of moral issues confronting investigators in research using human participants*. Jossey-Bass.
- Rienties et al. (2020). Defining the boundaries between artificial intelligence in education, computer-supported collaborative learning, educational data mining, and learning analytics: A need for coherence. *Frontiers in Education*, 5(128). <https://doi.org/10.3389/feduc.2020.00128>
- Rifkin, J. (2011). *La Tercera Revolución Industrial: cómo el poder lateral está transformando la energía, la economía y el mundo*. Paidós.
- Robert, J., & McCormack, M. (2024, 23 de mayo). 2024 EDUCAUSE action plan: AI policies and guidelines. EDUCAUSE Library. <https://library.educause.edu/resources/2024/5/2024-educause-action-plan-ai-policies-and-guidelines>
- Rojas, M.P., Chiappe, A. Artificial Intelligence and Digital Ecosystems in Education: A Review. *Tech Know Learn* **29**, 2153–2170 (2024). <https://doi.org/10.1007/s10758-024-09732-7>
- Romero, C., y Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- Romero, C., y Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Rosales Gallego, C., (2024). El principio de confianza aplicado al ámbito europeo de la inteligencia artificial. En *DACIU 2023/2024: Ciencias sociales y jurídicas* (pp. 116–124). Fundación Avanza. ISSN-e 3020-8173

- Roser, M., Ritchie, H., & Mathieu, E. (2023). *Technological change*. Our World in Data. <https://ourworldindata.org/technological-change>
- Ruder, S. (2018). An Overview of Multi-Task Learning in Deep Neural Networks. arXiv preprint. <https://doi.org/10.48550/arXiv.1706.05098>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Russell, S. J., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2 ed.). Upper Saddle River, NJ: Prentice Hall.
- Russell, S. y Norvig, P., (2021). *Inteligencia Artificial: Un enfoque moderno* (4 ed.). Madrid: Pearson.
- Sahota N., & Ashley, M. (2019). *Own the AI Revolution: Unlock Your Artificial Intelligence Strategy to Disrupt Your Competition*" (McGraw Hill)
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210–229. <https://doi.org/10.1147/rd.33.0210>
- Sanabria-Navarro, J.-R., Silveira-Pérez, Y., Pérez-Bravo, D.-D., y Cortina-Núñez, M. de-J. (2023). Incidencias de la inteligencia artificial en la educación contemporánea. *Comunicar*, 77, 97–107. <https://doi.org/10.3916/C77-2023-08>
- Sánchez Díaz, M. F. (2024). Inteligencia artificial generativa y los retos en la protección de los datos personales. *Estudios en Derecho a la Información*, 9(18), 179–205. <https://doi.org/10.22201/ijj.25940082e.2024.18.18852>
- Schmidgall, S., Su, Y., Wang, Z., Sun, X., Wu, J., Yu, X., Liu, J., Moor, M., Liu, Z., & Barsoum, E. (2025). *Agent Laboratory: Using LLM agents as research assistants* (arXiv:2501.04227). arXiv. <https://doi.org/10.48550/arXiv.2501.04227>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>

- Schwab, K. (2017). La cuarta revolución industrial. Ginebra: Foro Económico Mundial. Debate.
- Sedenberg, E., & Hoffmann, A. L. (2016). *Recovering the history of informed consent for data science and internet industry research ethics* (arXiv:1609.03266). arXiv. <https://doi.org/10.48550/arXiv.1609.03266>
- Siemens, G., & Baker, R. S. J. d 2012. Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12). Association for Computing Machinery, New York, NY, USA, 252–254. <https://doi.org/10.1145/2330601.2330661>
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. <https://doi.org/10.1177/0002764213479366>
- Smith, J. A. (2006). *Qualitative Psychology: A Practical Guide to Research Methods* (2 ed.). Londres: SAGE Publications.
- Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(7), 1880–1903. **[BUSCAR DOI]**
- Solove, D. (2010). *Understanding Privacy*. Harvard University Press. <https://www.hup.harvard.edu/books/9780674035072>
- Stein, E. (2005). *Sobre el problema de la empatía*. Trotta. (Obra original publicada en 1917).
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1410>
- Tamayo Velasco, B. J. (2021). Big data, competencia y protección de datos: el rol del reglamento general de protección de datos en los modelos de negocio basados en la publicidad personalizada. *Revista de Estudios Europeos*, 78, 183–202. <https://uvadoc.uva.es/handle/10324/48583>

- Tancredi, S., Abdu, R., Balasubramaniam, R., y Abrahamson, D. (2022). Intermodality in multimodal learning analytics for cognitive theory development: A case from embodied design for mathematics learning. En M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa y R. Hammad (Eds.), *The multimodal learning analytics handbook*. Springer.
- Tellez Dominguez , A. (2025). El derecho a la privacidad en México ante la IA y el big data: desafíos regulatorios. *Universita Ciencia*, 13(37). <https://dialnet.unirioja.es/servlet/articulo?codigo=10322995>
- Tene, O., & Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property*, 11(5). <https://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1>
- Tenorio Cueto, G. A. (2021). El derecho a una vida libre de algoritmos. *Revista Ius*, 15(48), 115–135. <https://doi.org/10.35487/rius.v15i48.2021.708>
- Tenorio Cueto, G. A. (Coord. ed.). (2019). *Ley Federal de Protección de Datos Personales en Posesión de los Particulares, comentada*. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI). ISBN 978-607-98648-1-1.
- Tenorio Cueto, G. A. (2026). Inteligencia Artificial, libertad de expresión y privacidad. La supremacía de la IA frente al deterioro de la libertad de expresión y la privacidad. Tirant Lo Blanch.
- Tomalá De La Cruz, M. A., Mascaró Benites, E. M., Carrasco Cachinelli, C. G., y Aroni Caicedo, E. V. (2023). Incidencias de la inteligencia artificial en la educación. *Recimundo*, 7(2), 238-251. [https://doi.org/10.26820/recimundo/7.\(2\).jun.2023.238-251](https://doi.org/10.26820/recimundo/7.(2).jun.2023.238-251)
- Tsai, Y.-S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., Drachsler, H., Delgado Kloos, C., & Gašević, D. (2020). Learning analytics in European higher education—Trends and barriers. *Computers and Education*, 155(May). <https://doi.org/10.1016/j.compedu.2020.103933>

- Tunzi, M., Satin, D. J., & Day, P. G. (2021). The Consent Continuum: A New Model of Consent, Assent, and Nondissent for Primary Care. *Hastings Center Report*, 51(2), 33–40. <https://doi.org/10.1002/hast.1243>
- Universidad Panamericana (2025). Marco general de uso ético de la IA. Ciudad de México: Universidad Panamericana.
- Uskov, V. L., Bakken, J. P., Howlett, R. J., & Jain, L. C. (Eds.). (2017). *Smart universities: concepts, systems, and technologies*. Springer.
- Van Alsenoy, B. (2016). Liability under EU data protection law: From Directive 95/46 to the General Data Protection Regulation. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law (JIPITEC)*, 7(3), 271–283. <https://www.jipitec.eu/issues/jipitec-7-3-2016/4539>
- Van Bekkum, M. (2025). Using sensitive data to de-bias AI systems: Article 10(5) of the EU AI Act. *Computer Law & Security Review*, 56, artículo 106115. <https://doi.org/10.1016/j.clsr.2025.106115>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30, 5998-6008. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act. *Computer Law Review International*, 22(4), 97–112. <https://doi.org/10.9785/cr-2021-220402>
- Vélez Rivera, R., Muñoz Álvarez, D., Leal-Orellana, P., & Ruiz-Garrido, A. (2024). Uso de inteligencia artificial en educación superior y sus implicancias éticas. Mapeo sistemático de literatura. Hachetetepe. *Revista científica de Educación y Comunicación*, 28, Art. 1105. <https://doi.org/10.25267/Hachetetepe.2024.i28.1105>
- Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., y Klerkx, J. (2014). Learning dashboards: An overview and future research

- opportunities. *Personal and Ubiquitous Computing*, 18(6), 1499–1514.
<https://doi.org/10.1007/s00779-013-0751-2>
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The Current Landscape of Learning Analytics in Higher Education. *Computers in Human Behavior*, 89, 98-110. <https://doi.org/10.1016/j.chb.2018.07.027>
- Voigt, P., & von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer.
<https://doi.org/10.1007/978-3-319-57959-7>
- Welch, V., Turner-Halliday, F., Watson, N., Wilson, P., Fitzpatrick, B., Cotmore, R., & Minnis, H. (2016). Randomisation before consent: avoiding delay to time-critical intervention and ensuring informed consent. *The International Journal of Social Research Methodology*. 1364-5579 (Print), 1464-5300.
<https://doi.org/10.1080/13645579.2016.1176751>
- Whitney, S. N., McGuire, A. L., & McCullough, L. B. (2004). A Typology of Shared Decision Making, Informed Consent, and Simple Consent. *Annals of Internal Medicine*, 140(1), 54–59. <https://doi.org/10.7326/0003-4819-140-1-200401060-00012>
- Williamson, B. (2018). *Big data en Educación. El futuro digital del aprendizaje, la política y la práctica*. Morata.
- Williamson, B., & Eynon, R. (2020). *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>
- Zarkadakis, G. (2015). *In Our Own Image: Will Artificial Intelligence Save or Destroy Us?* Londres: Rider.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>

Zeide, E. (2016). Student privacy principles for the age of big data: Moving beyond FERPA and FIPPs. *Drexel Law Review*, 8(2), 339–394.
<https://drexel.edu/law/lawreview/issues/Archives/v8-2/zeide/>

Documentos institucionales y estándares

Article 29 Working Party. (2018). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP251rev.01). European Commission.
https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

Comisión Europea (2020). Libro Blanco sobre la inteligencia artificial: Un enfoque europeo hacia la excelencia y la confianza. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX%3A52020DC0065>

Comité Europeo de Protección de Datos [CEPD]. (2020). *Directrices 5/2020 sobre el consentimiento en el sentido del Reglamento (UE) 2016/679* (Versión 1.1).
https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_es.pdf

European Data Protection Board. (2020). Guidelines 05/2020 on consent under Regulation 2016/679 (Version 1.1).
<https://www.edpb.europa.eu/sites/default/files/files/file1/edpbguidelines202005consenten.pdf>

Grupo de Expertos de Alto Nivel en Inteligencia Artificial. (2019). Ethics guidelines for trustworthy AI. Comisión Europea. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Holmes et al. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. UNESCO.
<https://unesdoc.unesco.org/ark:/48223/pf0000366994>

León XIV. (2026). *Magnifica humanitas* [Carta encíclica].
<https://www.vatican.va/content/leo-xiv/es/encyclicals/documents/20260515-magnifica-humanitas.html>

- Liu et al. (2023). Oportunidades y desafíos de la era de la inteligencia artificial para la educación superior: Una introducción para los actores de la educación superior. UNESCO. <https://doi.org/10.500.12592/nk98wdp>
- Naciones Unidas (1948). Declaración Universal de Derechos Humanos. <https://www.un.org/es/about-us/universal-declaration-of-human-rights>
- Naciones Unidas (1966). Pacto Internacional de Derechos Civiles y Políticos. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Department of Health and Human Services. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). <https://doi.org/10.6028/NIST.AI.100-1>
- Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (2018). The right to privacy in the digital age (A/HRC/39/29). <https://www.ohchr.org/Documents/Issues/DigitalAge/ReportPrivacyinDigitalAge/AHRC3929EN.pdf>
- Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (2021). The right to privacy in the digital age (A/HRC/48/31). <https://www.digitallibrary.un.org/record/3946475?v=pdf>
- Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). <https://oecd.ai/assets/files/OECD-LEGAL-0449-en.pdf>
- Organisation for Economic Co-operation and Development. (2024). Explanatory memorandum on the updated OECD definition of an AI system.

https://www.oecd.org/en/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en.html

Organización para la Cooperación y el Desarrollo Económicos. (2024). OECD AI Principles (actualización de mayo de 2024). OECD.AI Policy Observatory. <https://oecd.ai/en/ai-principles>

Organización de los Estados Americanos (1969). Convención Americana sobre Derechos Humanos (Pacto de San José). <https://www.oas.org/dil/esp/tratadosb-32convencionamericanasobrederechoshumanos.htm>

Organización para la Cooperación y el Desarrollo Económicos (OCDE) (2022). OECD Framework for the Classification of AI systems (OECD Digital Economy Papers, No. 323). OECD Publishing.

Pontificia Academia para la Vida. (2020). Rome Call for AI Ethics. <https://www.romecall.org/the-call/>

Red Iberoamericana de Protección de Datos [RIPD]. (2019). Orientaciones específicas para el cumplimiento de los principios y derechos que rigen la protección de los datos personales en los proyectos de inteligencia artificial. <https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-especificas-proteccion-datos-ia.pdf>

Secretaría de Educación Pública [SEP]. (2025). *Usos y percepciones sobre la inteligencia artificial generativa en la educación superior mexicana (ENIAG 2025): Dossier de resultados*. https://www.gob.mx/cms/uploads/attachment/file/1071411/Dossier_Resultados_ENIAG2025_SEP_.pdf

The Nuremberg Code. (1947). Trials of War Criminals Before the Nuremberg Military Tribunals Under Control Council Law No. 10 (Vol. 2, pp. 181–182). U.S. Government Printing Office. <https://history.nih.gov/display/history/Nuremberg+Code>

U. S. Department of Education. (s. f.). Protecting Student Privacy. <https://studentprivacy.ed.gov/>

- U.S. Department of Education, Office of Educational Technology. (2023). Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations. <https://www2.ed.gov/documents/ai-report/ai-report.pdf>
- UNESCO (2019). Beijing consensus on artificial intelligence and education. París: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
- UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- UNESCO. (2023). *Guidance for generative AI in education and research*. <https://doi.org/10.54675/EWZM9535>

Fuentes web y divulgación

- Anthropic. (2026). *Claude* [Modelo de lenguaje de gran escala]. <https://claude.ai/>
- Benemérita Universidad Autónoma de Puebla. (2026, 18 de febrero). Conformen el Consejo Académico Gobernanza de Inteligencia Artificial de la BUAP. Dirección General de Comunicación Institucional. <https://www.boletin.buap.mx/node/4287>
- Breceda Pérez, J. A. (2025). Hacia la regulación ética del uso de inteligencia artificial en la investigación científica: Lineamientos desde la Universidad Autónoma de Ciudad Juárez. *Revista de Bioética y Derecho*, 64, 81–97. <https://doi.org/10.1344/rbd2025.64.48719>
- Comisión Nacional de Bioética: México. (n. d.). Retrieved April 3, 2023, from <https://www.conbioetica-mexico.salud.gob.mx/interior/temasgeneral/consentimientoinformado.html>
- Coursera. (2025). *Supervised vs. unsupervised learning: Pros, cons, and when to choose*. <https://www.coursera.org/articles/supervised-vs-unsupervised-learning>
- Databricks. (s. f.). *What are AI agents?* Recuperado el 4 de agosto de 2026, de <https://www.databricks.com/glossary/ai-agents>

- Deloitte. (2024). Autonomous generative AI agents: Still under development. Deloitte Insights. <https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2025/autonomous-generative-ai-agents-still-under-development.html>
- Emory University, Office of Research Administration. (2023, 16 de septiembre). *Office of Research Administration to explore operations with cutting-edge generative AI and ChatGPT-like integration*. Emory Research Administration News. <https://scholarblogs.emory.edu/ranews/2023/09/16/office-of-research-administration-to-explore-operations-with-cutting-edge-generative-ai-and-chatgpt-like-integration/>
- Ferguson, R. (2012). The state of learning analytics in 2012: A review and future challenges (Technical Report KMI-12-01). Knowledge Media Institute, The Open University. <https://kmi.open.ac.uk/publications/techreport/kmi-12-01>
- Google. (2026). *Gemini* [Modelo de lenguaje de gran escala]. <https://gemini.google.com/>
- Higher Education Marketing. (2025). *The year of the AI agent in higher education*. <https://www.higher-education-marketing.com/blog/the-year-of-the-ai-agent-in-higher-education>
- Instructure. (2025, 23 de julio). Instructure and OpenAI announce global partnership to embed AI learning experiences within Canvas [Comunicado de prensa]. <https://www.instructure.com/press-release/instructure-and-openai-announce-global-partnership-embed-ai-learning-experiences>
- Instructure. (2026, 12 de marzo). Instructure delivers on its agentic AI promise with the launch of IgniteAI Agent [Comunicado de prensa]. <https://www.instructure.com/press-release/instructure-delivers-its-agentic-ai-promise-launch-igniteai-agent>
- International Association of Privacy Professionals (2024). State appeals court rules first CPRA regulations enforceable. <https://iapp.org/news/a/state-appeals-court-rules-cpra-regulations-can-take-effect>

- KeepCoding. (2024). *¿Qué es una red neuronal en Deep Learning?* KeepCoding Blog. <https://keepcoding.io/blog/red-neuronal-en-deep-learning/>
- Lama Muñoz (2025). Implicaciones jurídicas en materia de protección de datos de carácter personal como consecuencia de la utilización de la IA en la universidad. Artículos Científicos, 157. <https://sciencevalue.udit.es/articuloscientificos>
- Microsoft. (2025, julio). *Discover the potential of agentic AI in higher education.* Microsoft Education Blog. <https://www.microsoft.com/en-us/education/blog/2025/07/discover-the-potential-of-agentic-ai-in-higher-education/>
- Observatorio Interinstitucional de Inteligencia Artificial en la Educación Superior en México [OIIAES]. (2026). La incorporación de la Inteligencia Artificial en las instituciones de educación superior: primeros hallazgos del diagnóstico [Webinar]. <https://observatorio-ia.org>
- OpenAI. (2026). *ChatGPT* [Modelo de lenguaje de gran escala]. <https://chatgpt.com/>
- Perplexity AI. (2026). *Perplexity* [Motor de búsqueda con inteligencia artificial]. <https://www.perplexity.ai/>
- Sioli, L. (2021, 23 de abril). *An European strategy for artificial intelligence* [Presentación]. Comisión Europea. https://cdn.gelestatic.it/repubblica/blogautore/sites/408/2023/02/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21_compressed.pdf
- Soengas, M. (2025, April 29). Educación e IA: qué cambia con la nueva ley europea. <https://impulsoeducacio.org/es/educacion-e-inteligencia-artificial/>
- Tecnológico de Monterrey. (2024). Lineamientos para el uso ético de la inteligencia artificial. <https://tec.mx/es/integridad-academica/inteligencia-artificial>
- Universidad Autónoma de San Luis Potosí. (2026). Grupo Universitario de Inteligencia Artificial (GUIA). Secretaría Académica. <https://www.uaslp.mx/Secretaria-Academica/Paginas/Secretaria-Academica/9592>

Universidad Autónoma Metropolitana. (2026, 21 de abril). Crean decálogo ético para regular el uso de Inteligencia Artificial en la UAM. <https://boletines.uam.mx/archivos/crean-decalogo-etico-para-regular-el-uso-de-inteligencia-artificial-en-la-uam-numero237/>

Universidad Nacional Autónoma de México. (2026, 17 de marzo). Se crea Consejo Coordinador de Inteligencia Artificial. Gaceta UNAM. <https://www.gaceta.unam.mx/se-crea-consejo-coordinador-de-inteligencia-artificial/>

Universidad Panamericana. (2025). Marco general para el uso ético de sistemas de inteligencia artificial. Documento institucional.

Universitat Abat Oliba CEU. (2023). Plantilla de protocolo para el uso de la IA en actividades formativas. Vicerrectorado de Ordenación Académica y Calidad. https://www.uaoc.eu/sites/default/files/calidad/Protocolo_IA_Actividades_Formativas.pdf

WCET. (2026, 26 de marzo). *Agentic AI, academic integrity, and the question we are avoiding*. WCET Frontiers. <https://wcet.wiche.edu/frontiers/2026/03/26/agentic-ai-academic-integrity/>

Wilson Sonsini Goodrich & Rosati (2024). California appeals court moves up enforcement date for latest CCPA/CPRA regulations. <https://www.wsgr.com/en/insights/california-appeals-court-moves-up-enforcement-date-for-latest-ccpa-regulations.html>

World Compliance Association. (s. f.). *Sitio web*. <https://www.worldcomplianceassociation.com>

Legislación y normativa

California Privacy Protection Agency. (2018). California Consumer Privacy Act of 2018 (CCPA) (statute; versión consolidada publicada por la agencia, enero 2025). https://cppa.ca.gov/regulations/pdf/ccpa_statute.pdf

- European Data Protection Board & European Data Protection Supervisor. (2021). Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). <https://www.edpb.europa.eu/>
- European Data Protection Board & European Data Protection Supervisor. (2026). Joint Opinion 2/2026 on the proposal for a Regulation (Digital Omnibus). <https://www.edpb.europa.eu/>
- European Data Protection Board & European Data Protection Supervisor. (2026). Joint Opinion 2/2026 on the proposal for a Regulation (Digital Omnibus). <https://www.edpb.europa.eu/>
- FERPA. United States Congress. (1974). Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g. <https://www.law.cornell.edu/uscode/text/20/1232g>
- International Organization for Standardization. (2019a). ISO/IEC 27701:2019. Security techniques—Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management—Requirements and guidelines. ISO.
- International Organization for Standardization. (2019b). ISO/IEC 27018:2019. Information technology—Security techniques—Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors. ISO.
- International Organization for Standardization. (2020). ISO/IEC 29184:2020. Information technology—Online privacy notices and consent. ISO.
- International Organization for Standardization. (2023). ISO/IEC 42001:2023. Information technology—Artificial intelligence—Management system. ISO.
- International Organization for Standardization. (2023). ISO/IEC TS 27560:2023. Privacy technologies—Consent record information structure. ISO.
- LFPDPPP. Cámara de Diputados del H. Congreso de la Unión. (2025). Ley Federal de Protección de Datos Personales en Posesión de los Particulares (texto vigente). <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>

LGPDPESO. Cámara de Diputados del H. Congreso de la Unión. (2025). Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (texto vigente). <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGPDPEO.pdf>

Parlamento Europeo y Consejo de la Unión Europea (2016). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 (Reglamento General de Protección de Datos). <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>

Parlamento Europeo y Consejo de la Unión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de IA). <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

Parlamento Europeo y Consejo de la Unión Europea. (2026). Reglamento (UE) 2026/1744, de 8 de julio de 2026, por el que se modifican los Reglamentos (UE) 2024/1689, (UE) 2018/1139 y (UE) 2023/1230. Diario Oficial de la Unión Europea. <http://data.europa.eu/eli/reg/2026/1744/oj>

ANEXO 1

Con el propósito de asegurar la auditabilidad y la replicabilidad de la propuesta, el tablero de indicadores (KPI) presentado en el capítulo 4 se complementa con las fichas técnicas estandarizadas que se desarrollan a continuación. Estas fichas precisan, para cada indicador, su definición operativa, fórmula de cálculo, fuente de datos, periodicidad de medición, responsable, umbral/meta y riesgo asociado, permitiendo verificar la trazabilidad entre el MGC, las evidencias institucionales requeridas y el monitoreo continuo del cumplimiento.

Este anexo funciona como soporte metodológico y operativo para la adopción del modelo, al establecer criterios uniformes de medición y responsabilidades institucionales, sin alterar la lógica argumentativa del capítulo.

Anexo 1.1. Ficha técnica del tablero de indicadores clave de desempeño (KPI) del MGC.
(relacionada con la Tabla 10).

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
G 1	Gobernanza	Inventario de IA actualizado Mide el grado de actualización del registro institucional de sistemas de IA (herramientas propias y de terceros) frente a las herramientas efectivamente detectadas o en uso.	(# herramientas/casos registrados ÷ # detectados) ×100	Registro institucional de sistemas de IA; reportes de detección tecnológica.	Mensual	TD + CE-IA	≥ 95%	Uso de herramientas de IA no identificadas (“IA en la sombra”) y pérdida de la base única de rendición de cuentas.
G 2	Gobernanza	Clasificación de riesgo completada Mide el porcentaje del inventario de sistemas de IA clasificado conforme a la taxonomía mínima (prohibido, alto riesgo, riesgo limitado, riesgo mínimo).	(# clasificados ÷ total inventario) ×100	Inventario de casos; matriz de riesgo; dictámenes (Tabla 9).	Trimestral	CE-IA	100%	Aplicación de salvaguardas insuficientes a sistemas de alto riesgo por falta de clasificación previa.
G 3	Gobernanza	EIA/DPIA vigente en alto riesgo Mide la proporción de sistemas de alto riesgo con Evaluación de Impacto Algorítmico y/o DPIA vigente (≤12 meses).	(# alto riesgo con EIA/DPIA ≤12m ÷ total alto riesgo) ×100	Dictámenes de evaluación de impacto; matriz de riesgo (Tabla 9).	Trimestral	DPO + CE-IA	≥ 90%	Despliegue o continuidad de sistemas de alto riesgo sin evaluación vigente de impacto en derechos y datos personales.
G 4	Gobernanza	Cumplimiento del plan de auditorías	(# auditorías realizadas ÷ #	Informes de auditoría; bitácoras; métricas de	Semestral	CE-IA + TD	≥ 90%	Incumplimiento s no detectados oportunamente; pérdida de

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
		Mide el grado de ejecución del plan institucional de auditorías técnicas y de cumplimiento sobre IA y datos, frente a lo planificado.	planificadas) $\times 100$	desempeño (Tabla 9).				capacidad de auditoría interna/externa.
G 5	Gobernanza	Sesiones CE-IA realizadas Mide el funcionamiento efectivo del órgano colegiado, verificando la celebración de las sesiones planificadas del Comité de Ética en IA.	(# sesiones realizadas $\div$ # planificadas) $\times 100$	Actas; nombramientos; calendario; quórum; acuerdos (Tabla 9).	Trimestral	Secretaría técnica CE-IA	$\geq 90\%$	Debilitamiento del órgano de gobernanza y de la deliberación multidisciplinaria requerida para decisiones sobre IA.
G 6	Gobernanza	Compras de IA con evaluación Mide la proporción de adquisiciones de sistemas o herramientas de IA precedidas por la evaluación de términos, condiciones y cumplimiento normativo.	(# adquisiciones IA con evaluación previa $\div$ total adquisiciones IA) $\times 100$	Expedientes de compra (procurement); checklist de evaluación de proveedores.	Por evento	Compras + DPO + TD	100%	Contratación de proveedores que no garantizan protección de datos o soberanía digital.
G 7	Gobernanza	Acuerdos CE-IA cerrados con evidencia Mide el porcentaje de acuerdos del CE-IA en el periodo que cuentan con evidencia documental de cumplimiento o cierre.	(# acuerdos cerrados con evidencia $\div$ # acuerdos del período) $\times 100$	Actas; nombramientos; calendario; quórum; acuerdos (Tabla 9).	Trimestral	CE-IA	$\geq 80\%$	Acuerdos de gobernanza que quedan sin ejecutar, erosionando la autoridad del comité.
G 8	Gobernanza	Políticas críticas	(# políticas críticas)	Políticas publicadas; control de	Semestral	DPO + Jurídica	100%	Políticas desactualizadas frente a

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
		actualizadas en ≤12 meses Mide la vigencia de las políticas críticas (IA, datos, procurement), verificando su revisión o actualización dentro de los últimos 12 meses.	actualizadas en ≤12 m ÷ total políticas críticas) ×100	versiones; comunicados (Tabla 9).		+ CE-IA		cambios normativos o tecnológicos, con vacíos de gobernanza.
O 1	Operativa	Privacidad por defecto (training off cuando aplique) Mide la proporción de herramientas en las que el entrenamiento de modelos con datos de usuarios está desactivado por defecto, cuando existe esa opción.	(# herramientas con training OFF ÷ total con opción) ×100	Configuración enterprise; checklist técnico; validación técnica (Tabla 9).	Mensual	TD	100%	Uso no autorizado de datos del estudiantado/pr ofesorado para entrenar modelos de terceros.
O 2	Operativa	Fichas de transparencia publicadas Mide la cobertura de fichas técnicas de transparencia (capa 3 del consentimiento en capas) publicadas para las herramientas de IA activas.	(# herramientas con ficha ÷ total) ×100	Dashboard; fichas por herramienta; analíticas de consulta (Tabla 9).	Trimestral	TD + DPO	≥95%	Información insuficiente para un consentimiento realmente informado (incumplimiento del modelo de capas, 4.5.2.3).
O 3	Operativa	MTTR incidentes IA/datos Mide el tiempo promedio de resolución de incidentes relacionados con IA o datos, desde su apertura hasta	Promedio (días: apertura → cierre) por severidad	Bitácora / mesa de ayuda de incidentes IA-datos; informes de auditoría.	Mensual	TD + DPO	Críticos ≤15 d; Altos ≤30 d	Exposición prolongada de datos personales o fallas algorítmicas sin remediar.

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
		su cierre, por severidad.						
F 1	Formativa	Cobertura AI literacy docente Mide el porcentaje de docentes que aprobó la formación mínima institucional en alfabetización en IA (micro-credenciales).	(# docentes aprobados ÷ total docentes) ×100	LMS; constancias; evaluaciones pre/post (Tabla 9).	Semestral	Centro de formación docente	≥80%	Docentes sin competencias para mediar pedagógicamente el uso de IA, debilitando el consentimiento formado en el aula.
F 2	Formativa	Cobertura AI literacy estudiante Mide el porcentaje de estudiantes que aprobó la formación mínima institucional en alfabetización en IA.	(# estudiantes aprobados ÷ total estudiantes) ×100	LMS; constancias; evaluaciones pre/post (Tabla 9).	Semestral	Centro de formación docente + Vida estudiantil	≥70%	Estudiantado sin capacidad efectiva para ejercer un consentimiento informado ni cuestionar el uso de sus datos.
F 3	Formativa	Mejora pre/post en comprensión Mide el cambio en el nivel de comprensión sobre IA y protección de datos, comparando una evaluación diagnóstica previa y una posterior a la formación.	= promedio_post – promedio_pre (0–100)	Evaluaciones pre/post en LMS; constancias (mismo mecanismo que F1/F2).	Semestral	Evaluación institucional	≥ +15	Formación en IA sin impacto real en la comprensión; alfabetización meramente formal o simbólica.
P 1	Pedagógica	Cursos con declaración IA/datos Mide el porcentaje de cursos que declaran en su sílabo el uso de IA y el manejo de datos,	(# cursos con declaración ÷ # cursos que usan IA) ×100	Sílabos; guías; comunicación en LMS (Tabla 9).	Semestral	Coordinaciones + docentes	≥80%	Estudiantes que cursan asignaturas mediadas por IA sin haber sido informados desde el inicio del curso.

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
		conforme a la transparencia didáctica.						
P 2	Pedagógica	Pausas pedagógicas implementadas Mide la frecuencia con la que se implementan las pausas pedagógicas de reflexión crítica (“pausa triangular”) en cursos donde la IA interviene en evaluación o retroalimentación.	Promedio # pausas por curso/periodo	Planeación didáctica / bitácora docente; evidencias de sesión de clase.	Semestral	Docentes	$\geq 2/\text{curso}$	Ausencia de espacios de deliberación crítica sobre sesgos y límites de la IA en el aula.
P 3	Pedagógica	Revisión humana en decisiones de alto impacto Mide el porcentaje de decisiones de alto impacto (admisión, becas, permanencia, acreditación, sanciones) apoyadas en IA que fueron revisadas por una persona con capacidad de modificar el resultado.	$(\# \text{ decisiones alto impacto revisadas} \div \text{total}) \times 100$	Registro/expediente de decisiones; actas del CE-IA o del área responsable.	Trimestral	CE-IA + área responsable	100%	Decisiones automatizadas sin intervención humana significativa en procesos con efectos relevantes para el estudiantado.
R 1	Resultado	Índice de Confianza en IA (0–100) Mide, mediante instrumento estandarizado, el nivel de confianza institucional de estudiantes y docentes en el uso de IA en	Promedio encuesta (0–100)	Encuestas; grupos focales; reportes; observatorio (Tabla 9).	Semestral	Evaluación institucional / Observatorio	≥ 80	Erosión de la legitimidad institucional del uso de IA; desconexión entre lo normado y lo percibido.

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
		procesos académicos.						
R 2	Resultado	Índice de consentimiento informado (0–100) Mide, mediante instrumento estandarizado, el grado en que estudiantes y docentes perciben haber otorgado un consentimiento efectivamente informado.	Promedio encuesta (0–100)	Encuestas; grupos focales; reportes; observatorio (Tabla 9).	Semestral	DPO + Evaluación institucional / Observatorio	≥ 85	Consentimiento formalmente obtenido pero materialmente vacío (brecha documentada en ENIAG 2025, apdo. 3.6.3).
G 9	Gobernanza	Ejercicio efectivo de control por usuarios Mide la proporción de solicitudes de modificación o retiro de consentimiento presentadas por estudiantes o docentes que son atendidas sin condicionar, penalizar o restringir el acceso a servicios académicos, verificando que el control declarado en el modelo se traduzca en una capacidad real y libre de coerción.	(# solicitudes de modificación/ retiro de consentimiento atendidas sin penalización ÷ total de solicitudes recibidas) ×100	Registro de solicitudes de modificación o retiro; bitácora de atención de la DPO; portal de autoservicio de consentimiento (si existe).	Trimestral	DPO + CE-IA	100%	Consentimiento formalmente revocable pero materialmente inoperante: las personas usuarias enfrentan barreras, condicionamientos o represalias indirectas al ejercer su control, vaciando de contenido la garantía de libertad del consentimiento.
G 10	Gobernanza	Reclamaciones resueltas sobre decisiones de IA Mide la proporción de reclamaciones, impugnaciones o apelaciones presentadas por estudiantes o	(# reclamaciones o impugnaciones resueltas ÷ total de reclamaciones recibidas) ×100	Registro de reclamaciones; expedientes de resolución; actas del CE-IA o del área jurídica responsable.	Trimestral	CE-IA + Jurídica	≥ 90%	Decisiones automatizadas sin vía real de impugnación, dejando a las personas usuarias sin recurso frente a errores o resultados

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
		docentes respecto de decisiones o resultados derivados de procesos automatizados que son resueltas por la institución, verificando la existencia de una vía efectiva de reclamación frente al uso de IA.						adversos generados o asistidos por IA.

Nota. Elaboración propia con apoyo de Claude Sonnet 5 (Anthropic, 2026), El prompt fue el siguiente: "Revisa que el Anexo 1.1. (fichas técnicas de los KPI del MGC) sea correcto, completo, y enriquezca las Tablas 10 y 11. La redacción debe incluir KPI y definición operativa, fórmula/unidad, fuente de datos/evidencia, periodicidad, responsable, meta/umbral y riesgo asociado, a partir de las definiciones ya establecidas de control (capacidad efectiva de las personas usuarias para decidir y controlar el uso de la IA y de sus datos, sin coerción ni penalización) y reclamación (capacidad de impugnar, cuestionar o apelar decisiones o resultados asociados a la IA), e incorpóralas al documento conservando el formato de la tabla." El prompt identificó, mediante una revisión de correspondencia y completitud entre el Anexo 1a y las Tablas 10 y 11; las fichas fueron redactadas por la herramienta en el mismo formato y nivel de detalle que las diecinueve fichas preexistentes, a partir de las definiciones de control y reclamación desarrolladas por la autora, e incorporadas al documento previa revisión y validación de su contenido por parte de la autora.

Anexo 1.2 Ficha técnica de los KPI de consentimiento, confianza y auditabilidad (relacionada con la Tabla 11).

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
1	Gobernanza	Cobertura de evaluación de 'libertad' (no detrimento) en tratamientos basados en consentimiento Mide la proporción de tratamientos basados en consentimiento verificados como libres, es decir, sin condicionar servicios académicos esenciales ni generar perjuicio por negativa o retiro.	$(N \text{ tratamientos con checklist de libertad completo} / N \text{ total tratamientos basados en consentimiento}) \times 100$	Checklist de libertad del consentimiento; registro institucional de sistemas de IA.	Trimestral	CE-IA + DPO	≥ 95%	Consentimientos viciados por condicionalidades indebidas, contrario al estándar del CEPD (2020).
2	Operativa (técnico-operativa)	Granularidad del consentimiento por finalidad (desagregación) Mide el grado en que los flujos de datos separan sus finalidades (docencia, analítica, investigación) en solicitudes independientes de consentimiento.	$(N \text{ flujos con finalidades separadas} / N \text{ total flujos que requieren consentimiento}) \times 100$	Panel de preferencias (capa 4, Tabla 6); configuración técnica de la herramienta.	Semestral	TD + DPO	≥ 90%	Consentimiento “todo o nada” que impide autorizar unas finalidades y rechazar otras.
3	Gobernanza	Separación entre condiciones del servicio y consentimiento para tratamientos no necesarios Mide la proporción de flujos en que el consentimiento para tratamientos no esenciales se solicita separado de las condiciones generales de uso.	$(N \text{ flujos conformes} / N \text{ total flujos evaluados}) \times 100$	Registro institucional de sistemas de IA; auditoría de términos y condiciones.	Trimestral	DPO + Jurídica	100%	Consentimiento “empaquetado” que obliga a aceptar tratamientos no necesarios como condición del servicio.
4	Gobernanza + Pedagógica	Complejidad del paquete informativo en el momento de la solicitud Mide si, al solicitar el consentimiento, el usuario recibe la información mínima completa (aviso breve, finalidad, datos, opción real).	$(N \text{ flujos con paquete informativo completo} / N \text{ total flujos que requieren consentimiento}) \times 100$	Aviso breve y aviso ampliado (capas 1-2, Tabla 6); auditoría de interfaz.	Semestral	TD + DPO + Docentes	≥ 95%	Consentimiento solicitado sin información suficiente para una decisión informada.
5	Operativa (técnico-)	Acción afirmativa inequívoca (sin casillas premarcadas ni ambigüedad)	$(N \text{ flujos con opt-in inequívoco} / N \text{ total flujos})$	Checklist técnico de interfaz; configuración enterprise;	Trimestral	TD	100%	Consentimientos inválidos por diseño de interfaz que induce a la aceptación (dark patterns).

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
	operativa)	Mide la proporción de flujos que requieren una acción afirmativa clara (opt-in), sin casillas premarcadas ni patrones de diseño ambiguos.	evaluados) ×100	auditoría de diseño.				
6	Operativa (técnico-operativa) + Pedagógica	Simetría de retiro del consentimiento (es tan fácil retirar como otorgar) Mide la simetría entre los pasos requeridos para otorgar el consentimiento y los requeridos para retirarlo.	Índice = (Pasos para retirar / Pasos para otorgar)	Panel de preferencias; pruebas de usabilidad / auditoría técnica de interfaz.	Trimestral	TD	≤1.2	Fricción asimétrica que desincentiva el retiro del consentimiento.
7	Operativa (técnico-operativa)	Cumplimiento de plazos de retiro y cese del tratamiento Mide el porcentaje de solicitudes de retiro de consentimiento ejecutadas (cese efectivo del tratamiento) dentro del SLA institucional.	(N retiros ejecutados dentro del SLA / N retiros solicitados) ×100	Bitácora de solicitudes de retiro; mesa de ayuda; panel de preferencias.	Mensual	TD + DPO	≥95%	Continuidad del tratamiento pese a la revocación del consentimiento, sin base jurídica.
8	Gobernanza + Operativa (técnico-operativa)	Recambio/renovación de consentimiento ante cambios sustantivos Mide la proporción de cambios sustantivos en el tratamiento (nueva finalidad, proveedor o modelo) en que se obtuvo un nuevo consentimiento antes de aplicarlos.	(N cambios con re-consentimiento completado / N cambios sustantivos identificados) ×100	Registro institucional de sistemas de IA (control de versiones); panel de preferencias.	Trimestral	DPO + TD	≥95%	Tratamiento bajo finalidades distintas a las originalmente consentidas, sin renovar el consentimiento.
9	Operativa (técnico-operativa)	Trazabilidad del consentimiento (registro verificable) Mide la proporción de consentimientos otorgados o retirados con registro completo y verificable (quién, cuándo, finalidad, versión del aviso).	(N registros completos / N consentimientos registrados) ×100	Logs de auditoría; registro de consentimiento (ISO/TS 27560, 4.5.1.1).	Mensual	TD + DPO	≥99%	Imposibilidad de acreditar el consentimiento (carga de la prueba, art. 7.1 RGPD).
10	Gobernanza	Cobertura de Evaluación de Impacto (DPIA/AI-Impact) en procesos de alto impacto Mide la proporción de casos de uso de IA de alto impacto con evaluación de impacto (DPIA y/o EIA) vigente, conforme a la doble evaluación de 4.5.1.5.	(N casos IA alto impacto con evaluación vigente / N total casos IA alto impacto) ×100	Dictámenes de evaluación de impacto; matriz de riesgo (mismo mecanismo que G3).	Semestral	DPO + CE-IA	100%	Despliegue de sistemas de alto impacto sobre la trayectoria educativa sin evaluación previa de riesgos.
11	Gobernanza + Pedagógica	Notificación efectiva del uso de IA en decisiones significativas Mide el porcentaje de decisiones significativas	(N decisiones con notificación conforme / N total decisiones significativas con IA) ×100	Registro/expediente de decisiones; comunicación en LMS o notificación institucional.	Mensual	CE-IA + área responsable	≥95%	Decisiones que afectan la trayectoria del estudiantado sin que la persona sepa que hubo IA, impidiendo objetarlas.

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
		(admisión, becas, permanencia, acreditación, sanciones) en que la persona afectada fue notificada de la intervención de IA.						
1 2	Gobernanza + Operativa (técnico-operativa)	Revisión humana significativa (meaningful human oversight) disponible y usada Mide de forma combinada la disponibilidad de revisión humana en decisiones con IA y su uso efectivo mediante muestreo con evidencia.	$\text{Índice} = 0.5 \times (\% \text{ con opción de revisión}) + 0.5 \times (\% \text{ muestreo con evidencia de revisión})$	Registro/expediente de decisiones; actas del CE-IA; muestreo de auditoría (cf. P3).	Trimestral	CE-IA + área responsable	$\geq 90\%$	Revisión humana meramente nominal ("rubber-stamping"), sin capacidad real de modificar la decisión.
1 3	Gobernanza	Tasa de impugnación/resolución y tiempos de respuesta (garantías procedimentales) Mide la actividad del protocolo de objeción algorítmica: volumen de impugnaciones, tiempos de resolución y cumplimiento del SLA.	a) (Impugnaciones / Decisiones) $\times 1000$; b) promedio de días; c) (Resueltas en SLA/Impugnaciones) $\times 100$	Bitácora del protocolo de objeción algorítmica; registro de impugnaciones.	Trimestral	CE-IA + Jurídica	Definir por institución	Inexistencia o inoperancia de una vía real de impugnación frente a decisiones automatizadas.
1 4	Formativa + Pedagógica	Índice de comprensión del consentimiento (quiz breve post-aviso) Mide, mediante cuestionario breve aplicado tras el aviso, el grado real de comprensión de sus términos por el usuario (no solo la aceptación).	Puntaje promedio por unidad académica (0-100) + % que alcanza umbral ≥ 80	Quiz/instrumento de comprensión en LMS; reportes por unidad académica.	Semestral	Centro de formación docente + Evaluación institucional	Promedio ≥ 80	Consentimiento aceptado sin comprensión real ("no basta con que el usuario consienta; debe comprender", Sánchez Díaz, 2023).
1 5	Formativa	Cobertura de alfabetización en IA y privacidad para personal clave decisor y revisor Mide el porcentaje de personal con funciones de decisión o revisión sobre IA (CE-IA, DPO, TD, revisores humanos) con formación específica completada.	$(N \text{ personal clave capacitado} / N \text{ total personal clave}) \times 100$	LMS; constancias de formación (mismo mecanismo que F1).	Anual (seguimiento o trimestral)	Centro de formación docente + CE-IA	$\geq 90\%$	Personal con funciones críticas de supervisión sin capacidad técnica y ética para ejercerlas.
1 6	Gobernanza + Pedagógica + Formativa	Índice de confianza estudiantil en procesos con IA (instrumento estandarizado) Mide, mediante instrumento estandarizado, la confianza del estudiantado en procesos con IA, identificando la varianza (brecha) entre unidades académicas.	Promedio (0-100) + varianza entre unidades (brecha)	Encuestas; grupos focales; reportes; observatorio (mismo mecanismo que R1).	Anual (corte semestral si hay cambios)	Evaluación institucional / Observatorio	Definir por institución (benchmark)	Heterogeneidad no detectada entre unidades académicas, que puede ocultar problemas localizados.

#	Dimensión	KPI y definición operativa	Fórmula / unidad	Fuente de datos / evidencia	Periodicidad	Responsable	Meta / umbral	Riesgo asociado (si no se cumple)
17	Gobernanza	Tasa de quejas/reclamos vinculados a consentimiento e IA (por 10,000 usuarios) Mide la incidencia de quejas o reclamos relacionados con consentimiento o uso de IA, normalizada por cada 10,000 usuarios.	$(\text{Quejas en el periodo} / N \text{ usuarios}) \times 10,000$	Bitácora de quejas y reclamos; ruta de atención de quejas del registro institucional.	Trimestral	CE-IA + DPO	Tendencia a la baja	Incremento sostenido de quejas como señal temprana de fallas sistémicas en el consentimiento o la gobernanza.
18	Gobernanza + Operativa (técnico-operativa)	Cierre de acciones correctivas derivadas de auditorías de consentimiento/IA Mide la proporción de hallazgos de auditoría (consentimiento o IA) cerrados dentro del plazo comprometido, y su reincidencia.	$(\text{Hallazgos cerrados en plazo} / \text{Hallazgos totales}) \times 100; + \% \text{ reincidencia}$	Informes de auditoría; bitácoras de seguimiento de hallazgos (cf. G4).	Trimestral	CE-IA + TD	$\geq 90\%$ en plazo: reincidencia ↓	Hallazgos de auditoría que no se corrigen o se repiten, evidenciando gobernanza meramente declarativa.
19	Gobernanza + Operativa (técnico-operativa)	Uso correcto de consentimiento explícito cuando se tratan categorías especiales de datos Mide la proporción de flujos con categorías especiales de datos que cuentan con consentimiento explícito, en lugar de uno genérico.	$(N \text{ flujos con consentimiento explícito} / N \text{ total flujos con datos especiales}) \times 100$	Registro institucional de sistemas de IA; checklist de tratamiento de categorías especiales.	Semestral	DPO + TD	100%	Tratamiento de datos sensibles bajo consentimiento genérico e insuficiente.

Nota. Elaboración propia con apoyo de Claude Sonnet 5 (Anthropic, 2026), a partir del prompt "A partir de la Tabla 11 (KPI para implementar el MGC, relacionada con consentimiento, confianza y auditabilidad), genera las fichas técnicas estandarizadas de cada uno de los 19 indicadores. Para cada ficha incluye: número, dimensión del MGC a la que corresponde, el nombre del KPI seguido de su definición operativa en un párrafo que explique con precisión qué mide, la fórmula o unidad de cálculo, la fuente de datos o evidencia institucional que sustenta la medición, la periodicidad de medición, el responsable institucional, la meta o umbral sugerido, y el riesgo asociado en caso de que el indicador no se cumpla. Mantén el mismo formato de columnas y el mismo nivel de detalle en todas las fichas, de manera que el conjunto permita verificar la trazabilidad entre el modelo, las evidencias institucionales requeridas y el monitoreo continuo del cumplimiento" y del contenido conceptual y normativo desarrollado por la autora.