**Educaf5-Berit** ISSN L 3072-9971
Revista de Innovación y Reflexión Social en Educación

Vol.2, NÚM 1.
Septiembre 2026 a Agosto 2027
pp. 156-175

Artículo original

# Inteligencia artificial en educación superior: transformación sociotecnológica y desafíos éticos-regulatorios

Artificial Intelligence in Higher Education: Sociotechnological Transformation and Ethical-Regulatory Challenges

CARLOS GARCIA LOPEZ https://orcid.org/0009-0005-2632-3523

UNIVERSIDAD DE TECNOLOGIA Y EDUCACION UTED

carlos.garcial2024@uted.us

Recibido: 1 abril 2026
Aceptado: 29 junio 2026

URL: https://relaticpanama.org/_journals/index.php/educaf5-berit/article/view/171
DOI: https://doi.org/10.66707/dzc2ck91

## Resumen

Este ensayo reflexivo de corte documental examina la evolución histórica, los fundamentos técnicos y las implicaciones socioeducativas de la inteligencia artificial (IA), con énfasis en educación superior. Se parte de la trayectoria de la IA desde sus bases formales y hitos fundacionales hasta su consolidación contemporánea mediante aprendizaje automático, redes neuronales y aprendizaje profundo. A continuación, se analiza la incorporación de la IA en educación, confrontando la promesa de personalización y analítica del aprendizaje con críticas sobre tecnocratización, reducción del aprendizaje a métricas y riesgos para la integridad académica. Se discute el impacto social del uso generativo, destacando su

potencial como herramienta de ampliación cognitiva y, simultáneamente, sus riesgos epistemológicos y de sesgo. El texto integra marcos normativos y éticos, subrayando la centralidad humana y la supervisión significativa (UNESCO) y comparando enfoques regulatorios basados en riesgo (AI Act europeo) con desarrollos colombianos recientes (CONPES 4144 y Proyecto de Ley 043/2025–324/2025). Se concluye que la adopción universitaria de IA exige gobernanza institucional, alfabetización digital crítica, evaluación auténtica y mecanismos auditables de equidad, transparencia y responsabilidad, para que la innovación tecnológica fortalezca —y no sustituya— la mediación pedagógica y el sentido humanista de la educación superior.

**Palabras clave**
Inteligencia artificial, educación superior, ética y gobernanza, equidad algorítmica, inteligencia artificial generativa

## Abstract

This documentary-based reflective essay explores the historical development, technical foundations, and socio-educational implications of artificial intelligence (AI), with a focus on higher education. It reviews AI's evolution from its theoretical origins and foundational milestones to its current consolidation through machine learning, neural networks, and deep learning. The discussion examines the integration of AI into educational contexts, critically addressing the tension between the promise of personalized learning and data-driven decision-making and the risks of technocratic reductionism, metric-centered evaluation, and threats to academic integrity. The essay further analyzes the social impact of generative AI, emphasizing its dual character as both a tool for cognitive enhancement and a source of epistemological uncertainty and algorithmic bias. Ethical and regulatory frameworks are incorporated, highlighting human-centered principles and meaningful human oversight (UNESCO), while comparing risk-based regulatory approaches such as the European AI Act with recent Colombian policy developments (CONPES 4144 and Bill 043/2025–324/2025). The paper concludes that responsible integration of AI in higher education requires

institutional governance, critical digital literacy, authentic assessment strategies, and transparent accountability mechanisms to ensure that technological innovation reinforces—rather than displaces—the pedagogical and humanistic foundations of university education.

**Keywords**

Artificial intelligence, higher education, ethical governance, algorithmic equity, generative AI.

## Introducción

La inteligencia artificial (IA) se ha consolidado como uno de los avances tecnocientíficos más influyentes del siglo XXI, al posibilitar que sistemas computacionales realicen tareas asociadas tradicionalmente a la inteligencia humana —reconocimiento de patrones, aprendizaje y toma de decisiones— mediante técnicas como el aprendizaje automático y el aprendizaje profundo (Russell & Norvig, 2021; Goodfellow et al., 2016; LeCun et al., 2015). Sin embargo, su relevancia contemporánea no se agota en el incremento de potencia computacional, sino en la reconfiguración de prácticas sociales y educativas: desde plataformas adaptativas y analítica del aprendizaje hasta sistemas generativos que intervienen en la producción de conocimiento simbólico, alterando la autoría, la evaluación y las formas de mediación pedagógica (Holmes et al., 2019; Selwyn, 2019; UNESCO, 2023).

En educación superior, esta transformación introduce una tensión estructural entre promesas de personalización y riesgos de tecnocratización, en la medida en que la optimización algorítmica puede entrar en conflicto con el carácter ético, relacional y humanista de la formación universitaria (Luckin, 2018; Selwyn, 2019). A su vez, el despliegue de la IA acelera demandas de gobernanza responsable en torno a sesgos, equidad, transparencia, protección de datos y supervisión humana, principios enfatizados por la UNESCO (2021) y operacionalizados por marcos normativos basados en riesgo como el AI Act europeo (Reglamento [UE] 2024/1689). Al contextualizar en Colombia, esta agenda se

materializa tanto en política pública —CONPES 4144 de 2025— como en el trámite legislativo del Proyecto de Ley 043 de 2025 (Senado) – 324 de 2025 (Cámara), que incorpora obligaciones reforzadas para usos de IA en educación y formación cuando puedan afectar derechos.

En este marco, el presente ensayo reflexivo-documental analiza la trayectoria histórica y los fundamentos técnicos de la IA, confronta su impacto social —incluido el uso generativo— y argumenta la necesidad de una integración educativa regulada, ética y pedagógicamente fundada, orientada a fortalecer la mediación docente, la equidad y el desarrollo crítico del estudiante. Se describirán y analizarán sus fundamentos técnicos, su aplicación disruptiva, la repercusión social, las cuestiones ético - morales y la necesidad de regulaciones que garantizan su aplicación para beneficio de toda la humanidad, para finalizar, reflexionando sobre el futuro de la IA y sobre las vías para equilibrar la innovación y responsabilidad en esta era digital.

## Desarrollo

### Historia y desarrollo de la inteligencia artificial La inteligencia artificial (IA)

La inteligencia artificial (IA) posee antecedentes que se remontan mucho antes de la era digital. Sus raíces pueden rastrearse en los primeros intentos de la humanidad por comprender y replicar la naturaleza de la inteligencia. En la antigüedad, los mitos griegos ya hacían referencia a autómatas y seres artificiales, lo que evidencia una temprana fascinación por la creación de entidades capaces de actuar de manera autónoma.

Posteriormente, durante el siglo XIX, comenzaron a consolidarse los fundamentos formales que permitirían el desarrollo científico de la IA. Los avances en lógica y matemática fueron determinantes en este proceso. En particular, el trabajo de George Boole sobre la formalización del razonamiento lógico sentó bases estructurales para el desarrollo de sistemas computacionales capaces de operar mediante reglas formales.

La historia moderna de la inteligencia artificial inicia en la primera mitad del siglo XX, coincidiendo con el surgimiento y desarrollo de la computación. En 1943, Warren McCulloch

y Walter Pitts publicaron un modelo matemático de red neuronal artificial, considerado el primer intento formal de representar procesos neuronales mediante estructuras lógicas. Este aporte estableció los cimientos de lo que hoy se conoce como aprendizaje automático.

Más adelante, en 1950, Alan Turing propuso el célebre “Test de Turing” como un criterio para evaluar la capacidad de una máquina para exhibir un comportamiento inteligente indistinguible del humano. Esta propuesta no solo impulsó el debate científico sobre la posibilidad de emular la inteligencia humana, sino que también introdujo una discusión filosófica acerca de los límites entre mente, máquina y cognición.

En el mismo orden, es pertinente señalar que el concepto de "inteligencia artificial" surgió en 1956 como resultado de la Conferencia de Dartmouth, durante la cual investigadores como John McCarthy, Marvin Minsky y Allen Newell plasmaron la hoja de ruta de la investigación en inteligencia artificial. Durante las siguientes décadas, la inteligencia artificial vivió ciclos de grandes expectativas y diseño en lo que se conoce como " los inviernos de la inteligencia artificial ", pero también conoció desarrollos como la creación de sistemas expertos y el desarrollo de lenguajes de programación específicos como Lisp. La década de 1980 también es destaca por el auge de los sistemas expertos que ayudaban en tareas especializadas, por otro lado, en el siglo XXI la llegada del aprendizaje profundo y la explosión del Big Data han permitido renacer la inteligencia artificial y ayudarla a resolver problemas complejos en reconocimiento de voz, en visión por computadora o en procesamiento del lenguaje natural.

Visto así, la IA en la actualidad es una ciencia multidisciplinaria que integra matemáticas, estadística, informática y neurociencia, y está haciendo avances hacia sistemas más atractivos que necesariamente en una próxima evolución de la sociedad. Fundamentos técnicos de la IA La IA es un campo multidisciplinario que combina matemáticas, estadística, informática y neurociencia para construir sistemas que desarrollan tareas que requieren inteligencia humana. En el ámbito de la IA están los algoritmos, que se configuran como secuencias de instrucciones para que las máquinas puedan aprender y pueden tomar decisiones y resolver problemas a partir de datos.

- Algoritmos y aprendizaje automático El aprendizaje automático (o machine learning) es una de las subdisciplinas más relevantes del campo de la IA. Se apoya en la capacidad de los algoritmos para detectar patrones en los datos y mejorar su rendimiento con la experiencia, sin haber sido programados para hacer de cada tarea en particular. El aprendizaje automático suele dividirse en tres tipos:

Aprendizaje supervisado: El sistema aprende de un conjunto de datos etiquetados, donde se conoce la respuesta de cada entrada y que se utiliza en tareas, como clasificación y regresión, por ejemplo, para predecir el precio de una vivienda o identificar un correo electrónico único como spam.

Aprendizaje no supervisado: El sistema detecta patrones y relaciones en un conjunto de datos no etiquetados, descubriendo estructuras subyacentes, como grupos o asociaciones entre variables.

Aprendizaje por refuerzo: Un agente aprende por prueba y error a base de proporcionar recompensas o castigos en función de sus acciones, lo que es útil para la toma de decisiones en entornos dinámicos.

Desde esta perspectiva, la evolución histórica y técnica de la inteligencia artificial adquiere una relevancia particular en el ámbito educativo, donde sus desarrollos no solo representan innovaciones instrumentales, sino transformaciones estructurales en los procesos de enseñanza y aprendizaje. La incorporación de algoritmos de aprendizaje automático, sistemas adaptativos y análisis de grandes volúmenes de datos está redefiniendo la manera en que se diagnostican necesidades formativas, se personalizan trayectorias académicas y se toman decisiones pedagógicas basadas en evidencia.

En consecuencia, comprender los fundamentos técnicos y la trayectoria epistemológica de la IA no es un ejercicio meramente histórico, sino una condición necesaria para evaluar críticamente su potencial pedagógico, sus implicaciones didácticas y los desafíos éticos que plantea en la configuración de entornos educativos más inclusivos, eficientes y centrados en el estudiante.

Educaf5-Berit ISSN L 3072-9971
Revista de Innovación y Reflexión Social en Educación
Vol.2, NÚM 1.
Septiembre 2026 a Agosto 2027
pp. 156-175
Artículo original

**Redes neuronales y aprendizaje profundo**

Las redes neuronales artificiales constituyen modelos computacionales inspirados en la estructura y funcionamiento del cerebro humano. Aunque no reproducen fielmente los procesos biológicos, sí emulan la organización en capas de unidades interconectadas capaces de procesar información mediante pesos ajustables. En este sentido, estas arquitecturas permiten modelar relaciones no lineales complejas y aprender representaciones jerárquicas de los datos, lo que las convierte en herramientas particularmente eficaces para el reconocimiento de patrones (Goodfellow et al., 2016).

En este marco, el aprendizaje profundo (deep learning) representa una evolución significativa del aprendizaje automático tradicional, al apoyarse en redes neuronales profundas con múltiples capas que extraen progresivamente características de alto nivel a partir de grandes volúmenes de datos. Esta capacidad de aprendizaje jerárquico ha sido determinante en avances recientes en reconocimiento de voz, visión por computadora y procesamiento del lenguaje natural (LeCun et al., 2015). En efecto, el desempeño de sistemas contemporáneos en traducción automática, detección de imágenes o asistentes conversacionales se explica en buena medida por la consolidación de dichas arquitecturas.

Desde una perspectiva técnica, la inteligencia artificial se implementa en infraestructuras computacionales que integran hardware y software. Bajo esta lógica, la IA puede entenderse como el diseño de agentes racionales en sistemas computacionales capaces de percibir, procesar información y actuar en función de objetivos definidos (Russell & Norvig, 2021). Por su parte, la interacción humano-computadora (HCI) enfatiza que el diseño de interfaces influye de manera directa en la eficacia, la usabilidad y la dimensión ética de las aplicaciones basadas en IA (Shneiderman et al., 2016).

En síntesis, los fundamentos técnicos de la inteligencia artificial —redes neuronales profundas, aprendizaje automático y arquitecturas computacionales de alto rendimiento— han demostrado una capacidad extraordinaria para procesar información y resolver problemas complejos con niveles de precisión inéditos. Sin embargo, trasladar estos avances al ámbito educativo implica algo más que adoptar herramientas sofisticadas. Si bien

los sistemas basados en deep learning permiten personalizar trayectorias formativas, analizar patrones de desempeño y anticipar riesgos académicos, también plantean tensiones entre automatización y mediación pedagógica, eficiencia algorítmica y formación integral.

En el mismo orden, la educación no puede reducirse a un problema de optimización de datos; es un proceso humano, ético y relacional. Por ello, la incorporación de la inteligencia artificial en contextos educativos debe articularse con criterios pedagógicos sólidos, principios de equidad y responsabilidad social, de modo que la tecnología actúe como soporte del aprendizaje y no como sustituto de la intencionalidad formativa. En este sentido, el desafío no radica únicamente en desarrollar sistemas más inteligentes, sino en garantizar que su implementación contribuya al fortalecimiento de una educación crítica, inclusiva y centrada en el desarrollo humano.

**Incorporación de la inteligencia artificial al ámbito educativo**

La incorporación de la inteligencia artificial en educación ha sido presentada como un nuevo paradigma de personalización del aprendizaje, sustentado en sistemas adaptativos capaces de ajustar contenidos, ritmos y estrategias pedagógicas según el perfil y desempeño del estudiante. Holmes et al. (2019) sostienen que los sistemas basados en analítica de datos y aprendizaje automático permiten identificar patrones de progreso académico y ofrecer retroalimentación inmediata, fortaleciendo procesos de evaluación formativa y toma de decisiones pedagógicas basadas en evidencia.

No obstante, esta visión optimista debe ser matizada. Selwyn (2019) advierte que la promesa de personalización algorítmica puede invisibilizar dimensiones pedagógicas esenciales, al reducir el aprendizaje a métricas cuantificables y a trayectorias optimizadas por datos. Desde esta perspectiva crítica, la IA no es neutral, sino que incorpora supuestos epistemológicos y lógicas de mercado que pueden reconfigurar la función social de la

universidad. En consecuencia, el riesgo no radica únicamente en el sesgo técnico, sino en la posible tecnocratización de la experiencia educativa.

De manera complementaria, Luckin (2018) plantea que la inteligencia artificial solo aporta valor pedagógico cuando se integra dentro de un marco de “inteligencia aumentada”, en el que la tecnología amplía la capacidad del docente en lugar de sustituirla. Según esta autora, la clave no está en automatizar la enseñanza, sino en diseñar ecosistemas educativos donde la IA apoye la metacognición, el acompañamiento personalizado y la toma de decisiones informada.

Desde una perspectiva normativa, la UNESCO (2021) subraya que la IA en educación debe orientarse al fortalecimiento de la equidad, la inclusión y el acceso universal al conocimiento, evitando reproducir sesgos estructurales o ampliar brechas digitales. En este sentido, la personalización no puede entenderse únicamente como optimización algorítmica, sino como una herramienta para atender la diversidad, apoyar estudiantes en riesgo y promover trayectorias formativas flexibles y contextualizadas.

En el contexto colombiano, el debate regulatorio ha avanzado mediante proyectos legislativos que buscan establecer principios de transparencia, responsabilidad y protección de datos en el uso de sistemas inteligentes. El Proyecto de Ley sobre Inteligencia Artificial en trámite ante el Congreso de la República adopta un enfoque basado en riesgo y contempla obligaciones reforzadas cuando los sistemas puedan afectar derechos fundamentales, lo cual resulta particularmente relevante en educación superior, donde decisiones automatizadas podrían incidir en evaluación, permanencia o certificación académica.

En consecuencia, la integración de la IA en el sistema educativo colombiano no constituye únicamente un desafío técnico, sino un problema pedagógico, ético y jurídico. Su aplicabilidad debe articular innovación tecnológica con principios de gobernanza responsable y con una concepción humanista del aprendizaje, garantizando que los sistemas inteligentes fortalezcan el aprendizaje significativo, la mediación docente y el desarrollo crítico del

estudiante, en lugar de desplazar el componente relacional que define la experiencia universitaria.

En educación superior en Colombia, la discusión sobre IA ya no es solamente tecnológica: es, sobre todo, un problema de gobernanza académica y de garantía de derechos. En la práctica universitaria, los usos de IA (analítica del aprendizaje, tutores inteligentes, sistemas de recomendación, proctoring, chatbots institucionales, apoyo a docencia e investigación) reconfiguran decisiones sensibles: admisión y permanencia, evaluación del aprendizaje, certificación de competencias, bienestar estudiantil y gestión de datos. Por ello, la adopción "rápida" sin marcos de control tiende a producir tres tensiones: (a) eficiencia vs. equidad (sesgos y brechas digitales), (b) automatización vs. debido proceso (explicabilidad y posibilidad de reclamo), y (c) innovación vs. integridad académica (autoría, plagio, y evaluación auténtica). Esta lectura es consistente con la orientación de UNESCO hacia una visión "centrada en lo humano" y con exigencias de supervisión humana, transparencia y protección de derechos en el despliegue de IA en educación y educación superior.

En el plano colombiano, se observan dos capas complementarias. La primera es de política pública (transformación digital e IA), que enmarca prioridades de talento, infraestructura, datos y adopción sectorial. En esa línea se ubican los CONPES sobre transformación digital e IA, y más recientemente la Política Nacional de IA (CONPES 4144), que suele operar como brújula para programas, financiación y coordinación interinstitucional. La segunda capa es regulatoria/legislativa, visible en el trámite del Proyecto de Ley 043 de 2025 (Senado) – 324 de 2025 (Cámara), cuyo enfoque es explícitamente "basado en riesgo" y plantea gobernanza con autoridad nacional en cabeza de MinCiencias.

Para educación superior, lo más relevante es que el texto incorpora como "alto riesgo" a los sistemas de IA usados en educación y formación profesional que puedan determinar el acceso a la educación o la evaluación de estudiantes, precisamente por el potencial de impacto adverso sobre derechos. Además, exige evaluación de impacto previa para sistemas de alto riesgo, con componentes como análisis de riesgos (incluida discriminación

algorítmica), calidad/representatividad de datos, transparencia/explicabilidad y supervisión humana. En paralelo, el Ministerio de Educación ha señalado la construcción de una hoja de ruta para integrar IA en educación superior, lo que sugiere que la regulación "dura" (ley) y la conducción sectorial (MEN) tenderán a converger en estándares, orientaciones y mecanismos de aseguramiento de la calidad.

En el marco europeo, el punto de comparación obligado es el AI Act (Reglamento (UE) 2024/1689), porque fija un modelo regulatorio con obligaciones diferenciadas por niveles de riesgo y con consecuencias directas sobre aplicaciones educativas. En ese esquema, varios usos en educación y formación (p. ej., sistemas para evaluar resultados de aprendizaje, para orientar/condicionar trayectorias o acceso, y para monitoreo/detección de conductas prohibidas en exámenes como proctoring) aparecen tipificados como de "alto riesgo", lo que activa obligaciones reforzadas para proveedores y desplegadores.

Comparativamente, Colombia y la UE convergen en el "principio estructurante": la regulación por riesgo y la idea de que ciertas decisiones educativas mediadas por IA pueden afectar derechos y, por tanto, requieren cargas adicionales de diligencia (evaluación de impacto, transparencia y supervisión humana). Sin embargo, hay una diferencia práctica relevante para universidades colombianas: en Europa el AI Act opera como norma directamente aplicable y con un andamiaje de cumplimiento (conformity assessment, vigilancia de mercado, obligaciones detalladas), mientras que en Colombia el proyecto legislativo —aun en trámite— perfila principios, autoridad y mecanismos, pero su eficacia dependerá de reglamentación, capacidades institucionales, coordinación sectorial y producción de estándares técnicos (guías, protocolos, certificaciones) para que las IES puedan implementar cumplimiento verificable.

Implicaciones específicas para educación superior en Colombia:

1. Evaluación y decisiones académicas: todo uso de IA que incida en calificación, permanencia, admisión, alertas tempranas o "riesgo de deserción" debería tratarse como sistema potencialmente sensible (y, si calza en "alto riesgo",

exigir evaluación de impacto, trazabilidad y supervisión humana efectiva). Esto reduce arbitrariedades y fortalece el debido proceso académico.

2. Integridad académica y evaluación auténtica: la tendencia no debería ser solo “detectar IA”, sino rediseñar evaluación (oralidad, proyectos, desempeño situado) y exigir declaraciones de uso de IA, alineado con la orientación ética de UNESCO y con obligaciones de transparencia hacia estudiantes y docentes.

3. Datos y equidad: la promesa de personalización puede amplificar sesgos si los datos son incompletos o no representativos; por eso, la exigencia de analizar calidad/representatividad y mitigar riesgos no es un tecnicismo, sino una condición de justicia educativa.

4. Aseguramiento de la calidad: el diálogo anunciado por el MEN sugiere que la “hoja de ruta” puede conectarse con lineamientos institucionales (universidades/ASCUN y políticas internas) para traducir principios en prácticas auditables: gobernanza, comités de ética/IA, protocolos de compras (procurement), y formación docente.

**Impacto social del uso generativo y responsable de la inteligencia artificial**

La expansión de la inteligencia artificial generativa representa uno de los fenómenos sociotecnológicos más influyentes de la contemporaneidad, al introducir sistemas capaces de producir conocimiento simbólico —textos, imágenes, código, simulaciones— con niveles de coherencia y sofisticación sin precedentes. Desde una perspectiva tecnoptimista, Brynjolfsson y McAfee (2014) sostienen que estas tecnologías potencian la productividad cognitiva y amplían las capacidades humanas, configurando una “segunda era de las máquinas” caracterizada por la complementariedad entre inteligencia humana y automatización avanzada. En educación superior, este planteamiento sugiere que la IA generativa puede actuar como herramienta de ampliación cognitiva, facilitando la escritura académica, la modelación de problemas complejos y la tutoría personalizada.

No obstante, esta visión debe ser críticamente contrastada. Bender et al. (2021) advierten que los modelos generativos de gran escala operan mediante correlaciones estadísticas sin comprensión semántica genuina, lo que puede generar contenidos plausibles pero epistemológicamente frágiles, además de reproducir sesgos presentes en los datos de entrenamiento. Desde esta perspectiva, el impacto social de la IA generativa no es neutro: puede consolidar desigualdades informacionales, amplificar estereotipos y afectar la calidad del conocimiento producido en entornos académicos.

En una línea convergente, Selwyn (2019) sostiene que la digitalización educativa, incluida la IA, tiende a reconfigurar la educación bajo lógicas de eficiencia, medición y optimización, desplazando dimensiones críticas y humanistas del proceso formativo. En consecuencia, el riesgo no radica únicamente en la precisión técnica de los algoritmos, sino en la transformación cultural que estos inducen en las prácticas universitarias: evaluación automatizada, dependencia tecnológica y posible debilitamiento de la autoría intelectual.

Frente a estas tensiones, la UNESCO (2021) plantea que el desarrollo y uso de la inteligencia artificial deben fundamentarse en principios de centralidad humana, justicia, transparencia, rendición de cuentas y supervisión humana efectiva. La organización subraya que la IA debe fortalecer capacidades humanas y no sustituirlas, particularmente en sectores sensibles como la educación. Esta postura introduce un marco normativo que desplaza el debate desde la mera innovación tecnológica hacia la responsabilidad ética y la gobernanza institucional.

En el contexto latinoamericano, donde persisten brechas digitales y desigualdades estructurales, el impacto social de la IA generativa adquiere una dimensión adicional. La ausencia de políticas robustas de alfabetización digital crítica podría profundizar asimetrías entre estudiantes con acceso y competencias tecnológicas avanzadas y aquellos que carecen de ellas. Por tanto, el uso responsable de la IA generativa en educación superior no puede limitarse a regulaciones formales; requiere estrategias pedagógicas orientadas al desarrollo de pensamiento crítico, ética académica y competencias metacognitivas.

Educaf5-Berit ISSN L 3072-9971
Revista de Innovación y Reflexión Social en Educación
Vol.2, NÚM 1.
Septiembre 2026 a Agosto 2027
pp. 156-175
Artículo original

En síntesis, el impacto social de la IA generativa se sitúa en una tensión estructural entre ampliación cognitiva y riesgo de automatización acrítica del conocimiento. Su potencial transformador solo podrá consolidarse positivamente si se articula con marcos regulatorios claros, formación ética integral y una concepción humanista de la educación universitaria, en la cual la tecnología actúe como mediación y no como sustitución del ejercicio reflexivo y autónomo del sujeto.

**Sesgos algorítmicos y equidad**

Los sistemas de inteligencia artificial no operan en el vacío; se entrenan a partir de grandes volúmenes de datos históricos que reflejan dinámicas sociales preexistentes. En consecuencia, cuando dichos datos contienen desigualdades estructurales, los algoritmos pueden reproducirlas o incluso amplificarlas. O'Neil (2016) advierte que los modelos algorítmicos aplicados a decisiones sensibles —como selección de personal, crédito o admisión— pueden consolidar patrones discriminatorios al presentarse como neutrales y objetivos, cuando en realidad están moldeados por datos sesgados.

En una línea convergente, Barocas, Hardt y Narayanan (2019) explican que el sesgo algorítmico no siempre es producto de una intención explícita de discriminar, sino que puede emerger de variables correlacionadas con atributos sensibles como género, etnia o nivel socioeconómico. Esto implica que incluso sistemas técnicamente precisos pueden producir resultados injustos si no se evalúan críticamente sus supuestos y su desempeño diferencial entre grupos poblacionales.

El ámbito educativo no es ajeno a esta problemática. Sistemas de analítica del aprendizaje o modelos predictivos de deserción pueden clasificar a estudiantes como "de alto riesgo" con base en patrones históricos que reflejan desigualdades sociales estructurales. Sin mecanismos de supervisión humana y auditoría, estas herramientas podrían reforzar estigmatizaciones en lugar de ofrecer apoyos compensatorios.

Frente a este panorama, la UNESCO (2021) subraya que la equidad constituye un principio rector del desarrollo y uso de la inteligencia artificial. La organización enfatiza la

necesidad de incorporar evaluaciones de impacto ético, auditorías sistemáticas, diversidad en los equipos de desarrollo y mecanismos de transparencia y explicabilidad que permitan identificar y corregir sesgos a lo largo del ciclo de vida de los sistemas. La equidad algorítmica, por tanto, no es un atributo automático del diseño tecnológico, sino el resultado de una gobernanza deliberada y continua.

En consecuencia, el uso responsable de la IA exige un enfoque preventivo y correctivo. Esto implica establecer estándares de justicia algorítmica, implementar auditorías independientes, garantizar la supervisión humana efectiva y promover la participación interdisciplinaria en el desarrollo tecnológico. Solo bajo estas condiciones la inteligencia artificial podrá contribuir a la reducción de brechas y no a su profundización, especialmente en sectores sensibles como la educación superior.

**Autonomía, supervisión humana y gobernanza responsable**

El incremento en la capacidad de los sistemas de inteligencia artificial para analizar información, formular recomendaciones y ejecutar decisiones automatizadas ha reconfigurado la relación entre agencia humana y mediación tecnológica. A medida que los algoritmos asumen funciones tradicionalmente reservadas a la deliberación humana —como evaluación, clasificación o predicción— emergen interrogantes sobre la pérdida progresiva de control y la dilución de la responsabilidad. Floridi y Cowls (2019) sostienen que uno de los principios éticos fundamentales en la gobernanza de la IA es la preservación de la autonomía humana, entendida como la capacidad de los individuos para mantener control significativo sobre decisiones que afectan sus derechos y trayectorias vitales.

Desde una perspectiva complementaria, Mittelstadt et al. (2016) advierten que la automatización puede generar lo que denominan “desplazamiento de responsabilidad”, en el cual la toma de decisiones se atribuye al sistema algorítmico, dificultando la identificación de actores responsables. Este fenómeno resulta especialmente problemático en contextos educativos, donde decisiones automatizadas podrían influir en admisiones, evaluación académica o clasificación de riesgo estudiantil. La ausencia de supervisión humana efectiva

no solo compromete la equidad, sino también el debido proceso y el consentimiento informado.

En respuesta a estos desafíos, la UNESCO (2021) enfatiza el principio de supervisión humana significativa, estableciendo que los sistemas de IA deben diseñarse de manera que permitan intervención, revisión y corrección por parte de personas responsables. De manera convergente, el Reglamento Europeo de Inteligencia Artificial (AI Act, 2024) exige que los sistemas clasificados como de "alto riesgo" incorporen mecanismos de supervisión humana que eviten automatismos irreversibles en decisiones sensibles, particularmente en sectores como la educación y el empleo.

En consecuencia, el equilibrio entre autonomía tecnológica y control humano no debe entenderse como una oposición, sino como una relación de complementariedad regulada. La IA puede ampliar capacidades humanas, pero no debe sustituir la deliberación ética ni la responsabilidad institucional.

**Educación, políticas públicas y cultura de responsabilidad**

La velocidad del desarrollo tecnológico ha superado en muchos casos la capacidad regulatoria de los Estados, lo que ha generado una creciente demanda de marcos normativos claros que armonicen innovación y protección de derechos fundamentales. Jobin, Ienca y Vayena (2019), en un análisis comparativo de más de ochenta marcos éticos sobre IA, identifican convergencias globales en torno a principios como transparencia, justicia, no maleficencia y rendición de cuentas. Sin embargo, también señalan que la efectividad de dichos principios depende de su traducción en políticas públicas concretas y mecanismos de cumplimiento verificables.

En el contexto educativo, esta necesidad es particularmente apremiante. La implementación de IA en universidades exige políticas institucionales que regulen uso de datos, integridad académica, transparencia algorítmica y formación ética de docentes y estudiantes. La UNESCO (2023), en sus orientaciones sobre IA generativa en educación, subraya que la alfabetización digital crítica debe formar parte de las políticas educativas, de

modo que el uso de estas tecnologías no sea meramente instrumental, sino reflexivo y responsable.

Por tanto, el desafío contemporáneo no se limita a regular técnicamente la IA, sino a construir una cultura de desarrollo responsable que involucre a desarrolladores, instituciones educativas, empresas tecnológicas y gobiernos. Solo mediante marcos regulatorios claros, formación ética sistemática y supervisión institucional continua será posible maximizar los beneficios sociales de la inteligencia artificial y minimizar sus riesgos estructurales.

## Conclusiones

Primero, la IA no puede comprenderse únicamente como innovación técnica, sino como un fenómeno sociohistórico cuya madurez contemporánea se apoya en avances acumulativos de la lógica, la computación y los modelos de aprendizaje estadístico. Este recorrido explica por qué el salto reciente del aprendizaje profundo y la disponibilidad masiva de datos habilitan capacidades de predicción y generación que reconfiguran prácticas sociales, incluyendo las educativas (Goodfellow et al., 2016; LeCun et al., 2015; Russell & Norvig, 2021).

Segundo, en educación superior, la IA desplaza el debate desde “usar herramientas” hacia “redefinir condiciones formativas”. La personalización y la analítica del aprendizaje pueden fortalecer evaluación formativa y decisiones basadas en evidencia, pero, si se implementan sin criterios pedagógicos, pueden reducir el aprendizaje a indicadores y reforzar una racionalidad de optimización que tensiona la formación integral (Holmes et al., 2019; Selwyn, 2019). En esa línea, cobra fuerza la tesis de la “inteligencia aumentada”: la IA aporta valor cuando amplía la capacidad docente y la metacognición estudiantil, no cuando sustituye mediación pedagógica (Luckin, 2018).

Tercero, el uso generativo intensifica el dilema entre ampliación cognitiva y precarización epistemológica. Puede facilitar escritura académica y modelación de problemas, pero también producir contenidos plausibles sin garantías de verdad, reforzar sesgos y erosionar la autoría si no existen normas claras de transparencia y rediseño de

evaluación auténtica (Bender et al., 2021; Brynjolfsson & McAfee, 2014). La respuesta institucional no debería limitarse a "detectar IA", sino a fortalecer competencias críticas, prácticas de citación, trazabilidad del proceso intelectual y evaluaciones situadas.

Cuarto, la equidad algorítmica emerge como condición de justicia educativa. Los sistemas entrenados con datos históricos pueden reproducir desigualdades; por ello, el uso universitario de IA requiere auditorías, evaluación de impacto, revisión humana y gobernanza interdisciplinaria para evitar estigmatizaciones (Barocas et al., 2019; O'Neil, 2016; UNESCO, 2021). En educación, esto es decisivo en analítica de deserción, admisiones, proctoring y evaluación.

Quinto, la autonomía y el control humano deben considerarse criterios estructurantes. La automatización de decisiones educativas sensibles puede desplazar responsabilidad y debilitar el debido proceso; por tanto, la supervisión humana significativa y la rendición de cuentas deben ser exigibles y operables (Floridi & Cowls, 2019; Mittelstadt et al., 2016; UNESCO, 2021). Esta exigencia se alinea con enfoques regulatorios basados en riesgo, como el AI Act europeo (Reglamento [UE] 2024/1689).

Sexto, en Colombia la discusión se está traduciendo en política pública y debate legislativo: el CONPES 4144 fija capacidades y líneas de acción para la IA y el Proyecto de Ley 043/2025–324/2025 propone obligaciones reforzadas para sistemas de IA en educación y formación, especialmente cuando puedan afectar derechos, con foco en evaluación de impacto, transparencia y supervisión humana. En paralelo, el Ministerio de Educación Nacional impulsa un diálogo para construir una hoja de ruta de integración de IA en educación superior. En consecuencia, el reto inmediato para las IES no es solo adoptar IA, sino construir capacidades institucionales de cumplimiento, ética aplicada y aseguramiento de calidad, con políticas internas auditables y alfabetización digital crítica.

## Referencias

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. https://fairmlbook.org/

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton. Departamento Nacional de Planeación. (2025). *Política nacional de inteligencia artificial (Documento CONPES 4144)*. https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/4144.pdf

Congreso de la República de Colombia. (2025). *Proyecto de Ley 043 de 2025 (Senado) – 324 de 2025 (Cámara): "Por medio del cual se regula la inteligencia artificial en Colombia…"* (Informe de ponencia para primer debate). https://www.camara.gov.co/wp-content/uploads/2025/12/proyectos-ley/publicaciones/proyecto-35981/PPDSC-PL-043-25S-324-25C-INTELIGENCIA-ARTIFICIAL-SC.pdf

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign. https://curriculumredesign.org/our-work/artificial-intelligence-in-education/

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. UCL IOE Press.

Ministerio de Educación Nacional. (2025, 7 de marzo). *Educación superior inicia el camino para integrar la inteligencia artificial en el país* (Comunicado). https://www.mineducacion.gov.co/portal/salaprensa/Comunicados/423752%3AEducacion-superior-inicia-el-camino-para-integrar-la-inteligencia-artificial-en-el-pais

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2). https://doi.org/10.1177/2053951716679679

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

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 Inteligencia Artificial)*. *Diario Oficial de la Unión Europea*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the user interface: Strategies for effective human-computer interaction* (6th ed.). Pearson.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455

UNESCO. (2023). *Guidance for generative AI in education and research*. https://unesdoc.unesco.org/ark:/48223/pf0000386693

World Economic Forum. (2023). *The future of jobs report 2023*. https://www.weforum.org/reports/the-future-of-jobs-report-2023/