

Retos y oportunidades de la inteligencia artificial para la evaluación académica en la educación superior

Yamith José Fandiño Parra¹
Andrea Muñoz Barriga²



Resumen

La inteligencia artificial (IA) ha transformado las prácticas de evaluación en la educación superior, al redefinir cómo las instituciones miden el aprendizaje, proporcionan retroalimentación y apoyan el desarrollo académico. Este artículo

1. Doctorando, docente en la Facultad de Ciencias de la Educación de la Universidad de La Salle, Bogotá, Colombia. Cuenta con más de 20 años de experiencia docente en centros de idiomas y universidades. Se interesa por los estudios descoloniales, la formación del profesorado y la política lingüística. Correo electrónico: yfandino@unisalle.edu.co. ORCID: <https://orcid.org/0000-0002-5567-5465>
2. Doctora en Educación, docente en la Facultad de Ciencias de la Educación de la Universidad de La Salle, Bogotá, Colombia. Cuenta con más de 20 años de experiencia docente en universidades colombianas, con interés en la interculturalidad, la formación docente y la enseñanza de lenguas. Correo electrónico: munoz@unisalle.edu.co. ORCID: <https://orcid.org/0000-0002-0275-6528>

revisa la producción científica publicada entre los años 2024 y 2025, para reflexionar sobre las implicaciones pedagógicas de la evaluación impulsada por IA y destacar los avances emergentes y las contribuciones teóricas. La reflexión identifica cinco desafíos: sesgo algorítmico, riesgos de integridad académica, preocupaciones sobre la privacidad de los datos, la brecha digital y la posible deshumanización de la evaluación, junto con cinco oportunidades significativas, incluidas las vías de aprendizaje personalizadas, retroalimentación automatizada, análisis predictivo, mejora de la accesibilidad y mayor eficiencia en la evaluación. En este contexto, se propone un marco de integración ética y estratégica, que guíe al profesorado en la adopción de la IA como complemento, no como sustituto, del juicio humano. Los hallazgos sugieren que la IA fortalece la equidad educativa y la evaluación formativa si se implementa de forma transparente, responsable y con una guía institucional clara.

Palabras clave: inteligencia artificial, evaluación, evaluación formativa, educación superior.

Introducción

La integración de la inteligencia artificial (IA) en la educación superior constituye una innovación significativa que está transformando los procesos de enseñanza, aprendizaje y evaluación académica (Galán-Iñigo et al., 2025). Según Mpolomoka (2025), como profesores universitarios somos testigos de cómo la evaluación evoluciona de una medición estática de la retención del conocimiento a una herramienta dinámica, capaz de moldear experiencias de aprendizaje personalizadas mediante conocimientos basados en datos. Este cambio tecnológico es esencial para proporcionar retroalimentación oportuna a los estudiantes, evaluar su comprensión a profundidad y fomentar la mejora continua de sus trayectorias de aprendizaje.

En función de tales ideas, el objetivo de este artículo consiste en reflexionar sobre la producción científica y los avances teóricos relacionados con la IA en

la evaluación que se lleva a cabo en la educación superior. Se centró en el periodo del 2024 al 2025. Al revisar la literatura reciente, en este artículo se identificaron los posibles beneficios y limitaciones de la integración de la IA, al proporcionar una perspectiva equilibrada para los educadores que navegan esta disrupción (Cañar Torres *et al.*, 2025).

La literatura académica actual destaca que la aplicación de la IA en la evaluación incluye retroalimentación automatizada, predicción del rendimiento y el uso de *grandes modelos de lenguaje* (LLM, por su sigla en inglés), como ChatGPT (Díaz Pino, 2025). Estas herramientas tienen el potencial de mejorar la eficiencia y escalabilidad en entornos educativos, aunque también presentan desafíos complejos en cuanto a implementación y validez (Galán-Íñigo *et al.*, 2025). Es fundamental entender que la IA no solo automatiza prácticas existentes, sino que reconfigura la naturaleza de cómo verificamos la competencia y el rendimiento de los estudiantes. Por ello, este artículo explora las aplicaciones multifacéticas de la IA, que van desde el análisis diagnóstico hasta los dilemas éticos surgidos en contextos universitarios. En última instancia, el objetivo es proponer un marco que permita a los profesores aprovechar estas tecnologías de manera eficaz, al mantener, al mismo tiempo, el elemento humano esencial de la educación.

El alcance de esta reflexión abarca la transición de métodos tradicionales a estrategias innovadoras impulsadas por IA, que resaltan la necesidad de un enfoque fundamentado en la pedagogía en lugar de la tecnología de forma exclusiva (Biswas, 2025). En términos de Mpolomoka (2025), debemos considerar cómo estas herramientas afectan tanto al papel de estudiantes como de profesores, al asegurar que sirvan de apoyo y no como sustituto del juicio profesional. Esto requiere una revisión rigurosa de los riesgos asociados al sesgo algorítmico y de las oportunidades para crear entornos de aprendizaje más inclusivos. Al sintetizar los hallazgos de revisiones sistemáticas recientes y estudios empíricos, este artículo pretende ofrecer una visión general del

estado del arte. Las siguientes secciones detallarán el contexto, los desafíos y las oportunidades que definen esta nueva era de la evaluación educativa.

Tras haber esbozado la aparición de la IA en la evaluación de la educación superior y la motivación detrás de este artículo, es esencial situar la discusión en el contexto educativo y tecnológico actual. Comprender las condiciones actuales nos permite entender por qué las instituciones se ven presionadas para rediseñar los modelos de evaluación y cómo la IA ha acelerado esa necesidad.

Contexto y problema de la IA y la evaluación en la educación superior hoy

El panorama educativo actual se caracteriza por una explosión de datos y una necesidad urgente de contar con métodos de evaluación más eficientes y escalables para gestionar el crecimiento de la población estudiantil. En los últimos años, la rápida integración de herramientas de IA alteró paradigmas tradicionales, al obligar a las instituciones a replantearse cómo evalúan el rendimiento estudiantil y verifican los resultados de aprendizaje (Baltazar Flores, 2025). El problema radica en la tensión entre la eficiencia proporcionada por la automatización y la necesidad de preservar la integridad académica y el pensamiento crítico en el trabajo estudiantil. Según Fernández Cuevas (2025), aunque la IA ofrece soluciones para automatizar tareas repetitivas de calificación, en simultáneo introduce riesgos significativos respecto a la validez de las evaluaciones y al potencial de plagio.

Además, el rápido ritmo del avance tecnológico supera la capacidad de los marcos regulatorios y los programas de formación docente para mantenerse al día, cuando crea una brecha significativa en su implementación. En este sentido, Díaz Pino (2025) afirma que muchas instituciones en América Latina luchan contra una brecha digital, en la que el acceso desigual a la tecnología agrava las disparidades educativas existentes. Por su parte, Ávila Izquierdo *et al.* (2025) sostienen que la falta de políticas institucionales claras sobre el uso ético de la IA

generativa (GenAI) deja a muchos profesores inseguros sobre cómo incorporar estas herramientas de forma responsable en su plan de estudios.

Además, según Biswas (2025) y Mpolomoka (2025), preocupa, de forma generalizada, que depender demasiado de sistemas automatizados conduciría a la deshumanización del proceso educativo, al reducir a los estudiantes a simples datos. El asunto se agrava debido a la naturaleza de caja negra de muchos algoritmos de IA, que carecen de transparencia en cómo obtienen calificaciones o retroalimentación, lo que genera desconfianza entre los interesados. Los profesores enfrentan una mayor presión para detectar contenido generado por IA sin herramientas o formación adecuadas, lo cual conduce a una relación adversarial con la tecnología. Este entorno crea un dilema complejo: se sopesa el potencial de aprendizaje personalizado frente a los riesgos de vigilancia y pérdida de privacidad.

Como el panorama educativo actual atraviesa una profunda transformación impulsada por la rápida proliferación de herramientas de GenAI, se requiere una revisión crítica de los modelos de evaluación existentes en las universidades. Alario-Hoyos y Morales-Chan (2024) sostienen que, aunque la IA se usa desde hace años para personalizar contenido, la aparición de herramientas capaces de generar texto y código complejos exige una revisión completa de cómo las instituciones certifican el conocimiento y la autoría. Esta disrupción implica un periodo de incertidumbre y de adaptación de los profesionales de la evaluación, quienes ahora deben ir más allá de la eficiencia operativa para reimaginar la interacción entre tecnología y pedagogía (Miller, 2024). Además, sin un enfoque estratégico para desarrollar el pensamiento crítico, el sistema educativo corre el riesgo de volverse instrumental, cuando los estudiantes utilicen la IA como atajo en lugar de herramienta de desarrollo cognitivo (Flores Contrera, 2024).

La integración de estas tecnologías introduce desafíos éticos y logísticos complejos para los que las instituciones de educación superior a menudo no están preparadas. Kilińg expone que, aunque GenAI ofrece oportunidades

de asistencia personalizada, plantea preocupaciones significativas sobre la integridad académica y la interrupción de los modelos educativos tradicionales que solo dependen de la evaluación humana (2024). Existe una necesidad urgente de dominar estas herramientas emergentes para comprender sus implicaciones, no solo para la mecánica de la evaluación, sino para la experiencia educativa más amplia y la seguridad laboral de los educadores. También, la falta de políticas institucionales claras sobre su uso ético deja a muchos profesores inseguros sobre cómo incorporarlas de forma responsable, lo que crea una brecha de gobernanza (Hernández-León y Rodríguez-Conde, 2024).

Una vez establecido el panorama contemporáneo, la discusión se centra en las prácticas tradicionales de evaluación. Explorar métodos convencionales ayuda a poner de manifiesto las limitaciones persistentes y revela por qué los enfoques innovadores —en particular aquellos mediados por la IA— ganan terreno.

Desarrollo

Evaluación tradicional

La evaluación tradicional en la educación superior ha dependido históricamente de métodos estandarizados, como exámenes escritos, pruebas de opción múltiple y ensayos evaluados por los profesores a mano. Estos métodos, aunque establecidos, sufren sesgos inherentes, inconsistencias e importantes problemas de escalabilidad, en especial en grandes cursos de grado, en los que la atención individual es limitada. Llanos Torrico (2025) explica que el enfoque suele ser sumativo al medir qué sabe un estudiante en un momento concreto, en vez de diagnosticar su proceso de aprendizaje o potencial de crecimiento. Este enfoque de talla única no tiene en cuenta la diversidad de estilos de aprendizaje y las necesidades individuales de los estudiantes, lo que genera desigualdades en la forma de juzgar el rendimiento académico.

El bucle de retroalimentación en la evaluación tradicional suele retrasarse, al impedir que los estudiantes realicen correcciones inmediatas en su comprensión o estrategias durante el aprendizaje. De acuerdo con Mpolomoka (2025), la naturaleza laboriosa de la calificación manual limita la frecuencia y profundidad de la evaluación que los profesores proporcionan de forma realista, lo que a menudo resulta en comentarios superficiales. Factores como la fatiga del examinador, el sesgo inconsciente y la interpretación subjetiva influyen en la fiabilidad de las calificaciones asignadas en entornos tradicionales. Esta subjetividad supone una amenaza para la equidad, porque los estudiantes con niveles de rendimiento similares pueden recibir evaluaciones diferentes según quién califique su trabajo.

Los formatos tradicionales de evaluación, en adición, no logran captar competencias complejas como la creatividad, el pensamiento crítico y la resolución de problemas en contextos auténticos. Tienden a incentivar la memorización mecánica y el aprendizaje superficial antes que una comprensión conceptual profunda o la aplicación del conocimiento. Baltazar Flores (2025) sustenta que la estructura rígida de los exámenes no permite la flexibilidad para acomodar a estudiantes con diferentes orígenes o velocidades de aprendizaje. Esta falta de adaptabilidad ocasiona altos niveles de estrés y desinterés entre los estudiantes que sienten que la evaluación no refleja con precisión sus capacidades.

La mayoría de métodos tradicionales de evaluación en la educación superior se caracterizan históricamente por su dependencia de tareas puntuales y homogéneas que no reflejan la complejidad del aprendizaje. Según Alario-Hoyos y Morales-Chan (2024), estos métodos exigen bastantes recursos y ofrecen solo una visión limitada de las capacidades del estudiante en un instante específico, en lugar de promover una visión continua de su progreso. Al mismo tiempo, estas evaluaciones estandarizadas asignan las mismas tareas a todos los estudiantes, sin tener en cuenta sus diversos antecedentes o conocimientos previos, hasta limitar la capacidad de captar trayectorias de aprendizaje individuales. Flores Contrera (2024) añade que los exámenes tradicionales

no garantizan necesariamente un aprendizaje significativo; con frecuencia, no logran incentivar la capacidad de transformar el entorno o aplicar el conocimiento en contexto. En consecuencia, la rigidez de estos métodos es cada vez más incompatible con una era digital en la que la información es ubicua y la capacidad de procesarla críticamente es fundamental.

Por otra parte, los bucles de retroalimentación, inherentes a los modelos tradicionales de evaluación, suelen estar retrasados y carecen de la especificidad necesaria para impulsar una mejora inmediata. Miller (2024) establece que, aunque los profesionales tradicionales de la evaluación sobresalen en el mapeo y la elaboración de informes curriculares, la naturaleza estática de estas prácticas dificulta adaptarse a las necesidades dinámicas de los estudiantes modernos. En entornos tradicionales, la naturaleza laboriosa de la corrección manual limita la frecuencia de las evaluaciones e impide que los educadores ofrezcan la orientación que los estudiantes necesitan, en tiempo real, para corregir los malentendidos a medida que surgen. Esta desconexión hace que muchas evaluaciones tradicionales queden obsoletas, puesto que miden habilidades automatizables sobre las habilidades de pensamiento de orden superior, esenciales en una sociedad impulsada por IA.

Tras examinar las limitaciones de la evaluación tradicional, el siguiente paso es analizar cómo las alternativas impulsadas por IA transforman la evaluación. Esta sección compara modelos antiguos con nuevos, con el objeto de demostrar cómo la automatización, la personalización y la retroalimentación continua redefinen la medición del rendimiento y el aprendizaje en sí.

Evaluación innovadora

La evaluación innovadora impulsada por la IA introduce un cambio de paradigma hacia métodos formativos, continuos y adaptativos que priorizan el proceso de aprendizaje. Ávila Izquierdo et al. (2025) y Cañar Torres et al. (2025) sostienen que las tecnologías de IA crean vías de aprendizaje personalizadas, en las que

las evaluaciones se ajustan al nivel de competencia del estudiante y ofrecen una medida más precisa de sus capacidades. Los sistemas automatizados ahora procesan grandes cantidades de datos para brindar retroalimentación inmediata y detallada; característica que mejora de forma significativa la motivación y la autorregulación de los estudiantes. Las innovaciones incluyen el uso del procesamiento del lenguaje natural (PLN) para calificar tareas escritas complejas y la aplicación de análisis predictivo para identificar a los estudiantes en riesgo de quedarse atrás.

Además, la IA facilita la evaluación de resultados multimodales —imágenes, vídeos y presentaciones orales—, que amplían las formas en que los estudiantes demuestran su comprensión (Díaz Pino, 2025). Los sistemas de tutoría inteligente (STI) actúan como asistentes virtuales que diagnostican conceptos erróneos al instante y guían a los estudiantes a la comprensión correcta, pero sin intervención humana. Estas herramientas innovadoras proponen un modelo de evaluación continua, en que los datos se recogen de forma discreta durante las actividades de aprendizaje y reducen la ansiedad asociada a los exámenes de alto riesgo. De manera similar, Biswas (2025) asevera que, a través del análisis de patrones de aprendizaje, estos sistemas pueden generar informes completos, para que los profesores tengan una visión más profunda de la dinámica de clase y del progreso individual.

La implementación de estos métodos innovadores también promueve un entorno educativo más inclusivo, al adaptar el contenido a diversas necesidades, incluidas las de los estudiantes con discapacidad. Mpolomoka (2025) expone que los algoritmos adaptan la dificultad y el formato de las preguntas para ajustarse a la zona de desarrollo próximo del estudiante, en función de que las evaluaciones sean desafiantes, pero alcanzables. Este nivel de personalización era antes imposible de alcanzar a gran escala en la educación superior sin el apoyo de tecnología avanzada. En adición, la innovación impulsada por IA fomenta el desarrollo de habilidades de autoevaluación, cuando los estudiantes utilizan estas herramientas para revisar su propio trabajo y perfeccionarlo

antes de la entrega final. Así, la evaluación innovadora transforma el papel del estudiante: ya no es un receptor pasivo de calificaciones, sino un participante activo en su propia evaluación (Galán-Íñigo et al., 2025).

Según Avalos Guijarro (2024), la implementación de sistemas de IA en la evaluación reduce el tiempo de calificación en 40 %, mientras aumenta la consistencia de las evaluaciones en 30 %. Con ello, los educadores se centran en la retroalimentación cualitativa. Esta integración tecnológica, además, permite la creación de STI, el cual analiza los datos de rendimiento de los estudiantes, para ofrecer recomendaciones personalizadas e inmediatas y facilita el paso del juicio sumativo al apoyo formativo (Hernández-León y Rodríguez-Conde, 2024). Dönmez (2024) subraya que las herramientas de retroalimentación basadas en IA, en particular aquellas que utilizan PLN, pueden fomentar el aprendizaje autorregulado, al empoderar a los estudiantes a reflexionar sobre su rendimiento e identificar áreas de mejora de forma proactiva. Esta innovación transforma la evaluación en un socio de aprendizaje y no como un obstáculo, al promover una cultura que ve los errores como oportunidades de crecimiento. Mediante el uso de análisis de datos, estos sistemas predicen el rendimiento de los estudiantes y reconocen a las personas en riesgo de forma temprana, en función de asegurar que las intervenciones sean oportunas y efectivas.

Por otro lado, los marcos innovadores de evaluación se están ampliando para incluir dimensiones multimodales y colaborativas, las cuales, hace unos años, eran difíciles de gestionar a gran escala. Kılınc (2024) propone el *marco integral de evaluación de IA (MIEIA)*, el cual fomenta la integración ética de la IA en diversos niveles, desde la generación de ideas asistida por IA, hasta la colaboración completa en IA; con ello, la evaluación está alineada con las competencias digitales modernas. Este enfoque apoya la evaluación auténtica, en la que los estudiantes realizan tareas complejas y del mundo real, que pueden implicar el uso de la IA como copiloto, reflejando prácticas profesionales. Alario-Hoyos y Morales-Chan (2024) plantean que la innovación radica en evaluar el proceso y no solo el producto, mediante el uso de registros automáticos para rastrear

el progreso de los estudiantes y verificar la autoría en tiempo real. El uso de la IA también permite evaluar habilidades blandas y razonamientos complejos, mediante simulaciones y entornos interactivos. En estas áreas, los métodos tradicionales no son suficientes.

Aunque la evaluación asistida por IA innovadora ofrece ventajas sustanciales, la implementación no está exenta de obstáculos. Comprender los desafíos que amenazan la validez, la equidad y la integridad académica es clave para adoptar la IA de forma responsable en la educación superior.

Cinco desafíos de la IA respecto a la evaluación en la educación superior

El primer gran desafío es el sesgo *algorítmico*, porque los sistemas de IA entrenados con datos no representativos o históricos pueden reproducir o amplificar prejuicios sociales existentes. Según Díaz Pino (2025), esto propicia evaluaciones injustas de ciertos grupos demográficos estudiantiles, lo que perpetuaría la desigualdad en vez de resolverla (Dönmez, 2024). Este riesgo requiere vigilancia constante, ya que la lógica detrás de la calificación por IA es opaca y dificulta identificar la discriminación sin una auditoría rigurosa (Hernández-León y Rodríguez-Conde, 2024).

En segundo lugar, la *privacidad y ética de los datos* plantean preocupaciones importantes, porque la masiva recopilación de datos estudiantiles para analizar exige marcos sólidos de protección de información sensible y de garantía del consentimiento informado (Baltazar Flores, 2025). Las instituciones, por lo general, carecen de políticas claras necesarias para navegar estos campos minados éticos, lo que pone en riesgo los datos de los estudiantes para un mal uso. Por lo tanto, desarrollar marcos institucionales de gobernanza de datos se vuelve indispensable. Como señalan Alario-Hoyos y Morales-Chan (2024), el uso de herramientas externas de GenAI genera preocupaciones sobre la propiedad de los datos y derechos de autor de los estudiantes, lo que requiere políticas institucionales estrictas.

En términos de Fernández Cuevas (2025), un tercer desafío es la *integridad académica*. La proliferación de herramientas de GenAI dificulta distinguir entre el trabajo genuino de los estudiantes y el contenido generado por máquinas. Esta complicación hace que los métodos tradicionales de detección de plagio sean menos efectivos y obliga a replantearse cómo verificamos la autoría de los estudiantes. Flores Contrera (2024) lo ilustra al mostrar cómo las herramientas de IA pueden aprobar exámenes de educación superior con alta eficacia, lo que obliga a replantearse cómo verificamos la competencia estudiantil.

En cuarto lugar, existe una considerable *falta de formación docente*, pues muchos miembros del profesorado carecen de las competencias digitales necesarias para integrar y supervisar con eficacia estas herramientas avanzadas de evaluación (Galán-Íñigo et al., 2025). Sin la formación adecuada, los educadores harían mal uso de estas herramientas o no interpretarían correctamente sus resultados, lo que llevaría a tomar malas decisiones pedagógicas. Según Miller (2024), los profesionales de la evaluación deben desarrollar con urgencia nuevas habilidades para guiar al profesorado a través de este cambio tecnológico, con el objetivo de abrir una brecha cada vez mayor entre la disponibilidad de herramientas de IA y su aplicación pedagógica adecuada.

Por último, el *riesgo de deshumanización* en la educación es real. Depender demasiado de sistemas automatizados puede reducir la interacción humana esencial y el apoyo emocional que los profesores proporcionan. En palabras de Llanos Torrico (2025), ver a los estudiantes como datos y no como individuos erosiona la relación pedagógica, que es central en la educación superior. También existe una limitación técnica respecto a la precisión de la IA para evaluar tareas complejas, creativas o abiertas, para las cuales el juicio humano sigue siendo superior (Kılınç, 2024).

En contraste con los desafíos descritos, la IA también presenta oportunidades significativas. Destacar los beneficios favorece una toma de decisiones equilibrada

y fomenta el desarrollo de estrategias de evaluación que aprovechen la IA sin comprometer el propósito pedagógico.

Cinco oportunidades de la IA respecto a la evaluación en la educación superior

Una de las oportunidades más significativas es la *personalización del aprendizaje*. La IA analiza el rendimiento del alumnado para adaptar las evaluaciones y el contenido a las necesidades individuales, sus diferentes estilos y ritmos de aprendizaje (Díaz Pino, 2025). Esta capacidad brinda una experiencia educativa individualizada que antes era inalcanzable en grandes clases universitarias. Esto crea, en adición, vías de aprendizaje adaptativas donde la dificultad de las tareas se ajusta al progreso del estudiante de manera dinámica (Hernández-León y Rodríguez-Conde, 2024).

En segundo lugar, la IA ofrece una *eficiencia y automatización* inigualables, capaz de corregir grandes volúmenes de tareas al instante, lo que libera tiempo valioso dedicado a mentoría y formación compleja por parte de los docentes. Fuentes Cabrera et al. (2025) indican que esta reducción de la carga administrativa puede conducir a una mayor satisfacción del profesorado y a una mejor asignación de recursos. Por su parte, López Escobar et al. (2024) informan que las herramientas de IA mejoran significativamente la precisión de las evaluaciones, cuando reducen el tiempo y los recursos necesarios para su ejecución.

Según Baltazar Flores (2025), una tercera oportunidad es la *retroalimentación* en tiempo real y de alta calidad, para que los estudiantes corrijan errores de inmediato y se fomente un ciclo continuo de mejora que los métodos tradicionales no igualan. Esta inmediatez hace que la evaluación se convierta en una herramienta de aprendizaje poderosa, que anima a los estudiantes a involucrarse a profundidad con el material. Dönmez (2024) enfatiza, al mismo tiempo, que esta puntualidad es crucial en el aprendizaje autorregulado, al

ayudar a los estudiantes a supervisar su propio desarrollo académico de forma más eficaz.

En cuarto lugar, la IA propicia la *analítica predictiva* y la *intervención temprana*, para reconocer a los estudiantes en riesgo de abandono o fracaso al inicio del trimestre. Al mapear las tendencias de rendimiento estudiantil, las instituciones suelen ofrecer apoyo oportuno y mejorar las tasas de retención de forma significativa (Biswas, 2025). Además, dichos análisis proporcionan una comprensión más rica y matizada de los resultados de aprendizaje de los estudiantes, al incentivar intervenciones específicas que mejoran su retención (Miller, 2024).

Por último, la IA promueve la *inclusión* y la *accesibilidad*, con herramientas que adaptan las evaluaciones a estudiantes con discapacidades o barreras lingüísticas; hecho que democratiza una evaluación de calidad. Galán-Íñigo et al. (2025) afirman que estas herramientas traducen contenido, convierten texto a voz y ajustan la complejidad, para que todos los estudiantes tengan la misma oportunidad de demostrar sus conocimientos. En adición, la IA apoya el análisis de evaluaciones multimodales, que no solo se enfoca en la evaluación de diversos tipos de trabajos estudiantiles más allá del texto, sino que brinda opciones que posicionan a la IA como un aliado poderoso para mejorar los resultados y la equidad. En la tabla 1, se sintetizan los cinco desafíos y las cinco oportunidades discutidas hasta el momento.

Tabla 1. Desafíos y oportunidades de la IA respecto a la evaluación en la educación superior

Categoría	Síntesis	Elementos clave
Desafío 1: sesgo algorítmico	Riesgo de reproducir desigualdades sociales y calificaciones injustas	- Datos no representativos - Opacidad del algoritmo - Necesidad de auditorías constantes

Categoría	Síntesis	Elementos clave
Desafío 2: privacidad y ética de datos	Riesgos de manejo indebido de información estudiantil	• Falta de políticas institucionales claras • Problemas de consentimiento • Propiedad de datos
Desafío 3: integridad académica	Dificultad para distinguir autoría real frente a contenido generado por IA	• Ineficacia de métodos tradicionales • IA capaz de aprobar exámenes • Replanteamiento de verificación de competencias
Desafío 4: falta de formación docente	Docentes sin competencias para integrar y supervisar la IA	• Riesgo de mal uso • Interpretaciones erróneas • Necesidad de capacitación continua
Desafío 5: deshumanización del proceso	Reducción de la interacción pedagógica y afectiva	• Riesgo de ver al estudiante como dato • Limitaciones de la IA en tareas creativas y abiertas
Oportunidad 1: personalización del aprendizaje	Adaptación dinámica a estilos, ritmos y necesidades de cada estudiante	• Vías adaptativas • Ajustes de dificultad • Análisis continuo del rendimiento
Oportunidad 2: eficiencia y automatización	Reducción significativa del tiempo de calificación y carga administrativa	• Corrección automática • Ahorro de tiempo • Mayor dedicación a mentoría
Oportunidad 3: retroalimentación inmediata	Mejora del aprendizaje autorregulado mediante comentarios en tiempo real	• Correcciones instantáneas • Mayor motivación • Ciclos formativos continuos
Oportunidad 4: analítica predictiva	Identificación temprana de estudiantes en riesgo	• Predicciones basadas en patrones • Mejora de retención • Intervenciones oportunas
Oportunidad 5: inclusión y accesibilidad	Evaluaciones más flexibles y adaptadas a necesidades diversas	• Traducción • Simplificación • Multimodalidad • Apoyo a estudiantes con discapacidades

Fuente: los autores.

Al plantear los desafíos y las oportunidades, las universidades necesitan orientación práctica para regular y operacionalizar la evaluación mediada por IA. Por ello, esta sección propone marcos, para que los docentes integren la IA de forma ética, estratégica y sostenible.

Marco de incorporación de la IA respecto a la evaluación en la educación superior

Para incorporar la IA con eficacia, Mpolomoka (2025) aconseja a los profesores adoptar el marco de conocimiento tecnológico de contenidos pedagógicos (TPACK, por sus siglas en inglés), en función de cómo integrar la tecnología con su materia específica y estrategias de enseñanza. No solo basta con usar herramientas de IA, deben integrarlas de manera que realce los objetivos pedagógicos del curso. Es esencial mantener un enfoque de *humano en el bucle*, porque la IA sugiere calificaciones o retroalimentación, pero el juicio final y la interpretación contextual son responsabilidad del educador. Esta supervisión garantiza que se respeten las sutilezas del trabajo del alumnado y que se corrijan los errores algorítmicos.

Los profesores deberían centrarse en diseñar evaluaciones *resistentes a la IA o colaborativas con IA*, como defensas orales, análisis críticos de los resultados de la IA o aprendizaje basado en proyectos que requiera una visión humana única. Este cambio implica alejarse de tareas simples de recuerdo hacia pruebas que evalúen habilidades de pensamiento de orden superior, en las que la IA tiene dificultades para competir. Desarrollar directrices éticas claras y *prompts* es puntual, puesto que el profesorado debe enseñar a los estudiantes cómo utilizar las herramientas de IA de forma responsable, explícita y transparente en sus tareas. Esto exige crear políticas de divulgación escritas, informes reflexivos o el uso guiado de LLM respecto a tareas académicas específicas (Galán-Íñigo et al., 2025).

En este contexto, el desarrollo profesional continuo es clave para mantenerse al día con los rápidos avances de la IA, desde la alfabetización básica a la aplicación avanzada en el diseño curricular. En palabras de Baltazar Flores (2025), las instituciones deben apoyar este desarrollo al brindar recursos y formación que permitan al profesorado auditar y seleccionar herramientas de IA adecuadas. Por último, estos últimos deben utilizar la analítica de IA no solo para calificar, sino también para la enseñanza reflexiva, con el uso de datos para ajustar sus propios métodos de enseñanza y así servir mejor a la clase.

Para incorporar eficazmente la IA en la evaluación, los profesores tienen que adoptar un marco estructurado que haga hincapié en la alfabetización ética, la alineación pedagógica y la adaptación continua. Kılınç (2024) introduce el MIEIA, el cual guía a los docentes a través de niveles específicos de integración de IA, que van desde *sin IA* hasta *IA completa*, con la finalidad de que el nivel de uso de la tecnología sea adecuado para los objetivos de aprendizaje y la etapa educativa de los estudiantes. Este marco exige que los profesores enseñen a los estudiantes los principios éticos de transparencia y rendición de cuentas, al obligarles a revelar el grado de asistencia en IA en su trabajo. Miller (2024) refuerza esta idea cuando argumenta que los profesionales de la evaluación necesitan evolucionar hasta convertirse en *líderes en IA*, en colaboración con el resto del profesorado, para diseñar evaluaciones resistentes a la manipulación de la IA o que la incorporen de forma productiva para mejorar el pensamiento crítico. Esta transición implica desarrollar la creatividad digital y saber evaluar la participación de la IA en la producción académica. Al establecer directrices claras, los profesores fomentarían un entorno en el que la IA sea una herramienta de empoderamiento en lugar de una amenaza para la integridad.

Además de la estructura ética, este marco debe priorizar el enfoque de las personas en el bucle para garantizar una evaluación válida y significativa. Hernández-León y Rodríguez-Conde (2024) exponen que, aunque la IA automatiza el análisis de datos y el reconocimiento de patrones, la interpretación final de estas métricas aún es responsabilidad humana, para tener en cuenta

factores contextuales y emocionales que los algoritmos no perciben. Los profesores tendrían que utilizar análisis impulsados por IA para realizar una enseñanza reflexiva, con el uso de los datos de rendimiento estudiantil para ajustar sus estrategias de enseñanza en tiempo real (Dönmez, 2024).

En una línea similar, Alario-Hoyos y Morales-Chan (2024) sugieren que un marco sólido también implica rediseñar la evaluación para centrarse en el proceso de creación, mediante defensas orales y verificación en clase que complementen los resultados generados por IA. Esto exige un modelo dual de transformación digital: la infraestructura técnica apoya, en vez de dictar, los objetivos pedagógicos. En última instancia, el marco empodera a los profesores a actuar como mentores que guían a los estudiantes en la cocreación responsable del conocimiento con la IA. La figura 1 ofrece una visión general del marco propuesto.

Figura 1. Marco para incorporar la IA en la evaluación en educación superior



Fuente: los autores.

Conclusión

La integración de la IA, en la evaluación de la educación superior, es una tendencia irreversible con oportunidades profundas para mejorar la calidad y equidad del aprendizaje. Aunque la IA proporciona herramientas de eficiencia, personalización y visión analítica, también requiere un enfoque vigilante ante desafíos éticos como el sesgo, la privacidad y la integridad académica. La transición de la evaluación tradicional a la innovadora demanda no solo inversión tecnológica, sino un cambio cultural dentro de las universidades para valorar la pedagogía basada en datos junto con la mentoría humana. Los profesores desempeñan un papel primordial en este ecosistema, al actuar no como entidades reemplazables, sino como guías expertos que aprovechan la IA para fomentar el pensamiento crítico y la creatividad.

El éxito futuro depende de establecer marcos regulatorios sólidos y asegurar que la tecnología esté subordinada a los objetivos pedagógicos y los valores humanos. La evidencia sugiere que, cuando se implementa con transparencia y formación adecuada, la IA puede mejorar significativamente el rendimiento académico y la satisfacción del alumnado. En última instancia, una implementación equilibrada y ética de la IA en la evaluación será decisiva para definir el futuro de la educación superior en Colombia y el mundo. Esta sinergia entre la inteligencia humana y la eficiencia artificial es la clave para contar con un entorno educativo más inclusivo y eficaz.

Declaración ética sobre el uso de herramientas de GenAI

Los autores declaramos que, en el proceso de elaboración, revisión y edición de este artículo, empleamos de manera estratégica herramientas de GenAI, como Perplexity para la consulta, organización y verificación preliminar de información, y Quillbot para apoyar tareas lingüísticas de reescritura y mejora de estilo. Dichas herramientas se utilizaron como recursos complementarios, sin sustituir en ningún momento el juicio académico, la fundamentación teórica ni

el análisis crítico de los autores. Nos responsabilizamos íntegramente del contenido, la interpretación de los datos y las conclusiones presentadas. Todas las fuentes bibliográficas incorporadas se verificaron a mano, se citaron conforme a las normas académicas vigentes y se sometieron a revisión editorial independiente. El uso de IA se circunscribió a la optimización de la redacción y no a la generación autónoma de aportes intelectuales originales.

Referencias

Alario-Hoyos, C. y Morales-Chan, M. (2024). Evaluación educativa en tiempos de la inteligencia artificial. En M. T. Arteaga Auquilla y J. Maldonado Mahauad (Eds.), *Transformación de la evaluación en la educación superior, desde un enfoque estratégico y reflexivo* (pp. 63-72). Universidad de Cuenca.

Avalos Guijarro, A. Á. (2024). Impacto de la inteligencia artificial en la evaluación y retroalimentación educativa. *Retos para la Investigación*, 3(1), 19-32. <https://doi.org/10.62465/rri.v3n1.2024.72>

Ávila Izquierdo, P. E., Murillo Piza, V. A., Muyulema Cadena, S. L. y Navarro Pérez, Y. M. (2025). La inteligencia artificial como apoyo para mejorar la evaluación educativa. *Sinergia Académica*, 8(2), 217-233. <https://sinergiaacademica.com/index.php/sa/article/view/553>

Baltazar Flores, R. A. (2025). Aplicaciones de la inteligencia artificial en la evaluación del aprendizaje en la educación superior: beneficios, limitaciones y desafíos éticos. *Innovarium International Multidisciplinary Journal*, 3(2), 1-13. <https://tinyurl.com/26qpn5hd>

Biswas, S. (2025). AI and assessments & evaluations in education. In B. Jana, P. Sarkar, P. Lepcha y P. Topno (Eds.), *Pedagogy to cybergogy - education 5.0: Promises, theories, methods and designs* (pp. 96-105). LAMBERT Academic Publishing.

- Cañar Torres, M. S., Faican Quinche, E. J., Ninahualpa Aguiar, A. E., Criollo Peralta, D. J., Caiza Hidalgo, L. E. y Estacio Moreno, J. S. (2025). Evaluación automatizada mediante inteligencia artificial: beneficios y limitaciones. *South Florida Journal of Development*, 6(8), e5725. <https://doi.org/10.46932/sfjdv6n8-042>
- Díaz Pino, A. A. (2025). Inteligencia artificial y evaluación en contextos universitarios: una revisión de los últimos avances. *Educación y Futuro*, (52), 67-83.
- Dönmez, M. (2024). AI-based feedback tools in education: A comprehensive bibliometric analysis study. *International Journal of Assessment Tools in Education*, 11(4), 622-646. <https://doi.org/10.21449/ijate.1467476>
- Fernández Cuevas, H. (2025). Evaluación formativa e inteligencia artificial: estrategias para un aprendizaje humano y eficaz. *MLS Pedagogy, Culture and Innovation*, 2(1), 126-141. <https://tinyurl.com/297vdzar>
- Flores Contrera, C. J. (2024). La evaluación educativa en la era de la inteligencia artificial; cambios de paradigmas. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(1), 1579-1591. <https://doi.org/10.56712/latam.v5i1.1694>
- Fuentes Cabrera, C. J., Tapia Zurita, O. G. y Tapia Herrera, D. V. (2025). Uso de la inteligencia artificial en la evaluación formativa y su incidencia en el desempeño académico de los estudiantes. *Revista Social Fronteriza*, 5(1), e611. [https://doi.org/10.59814/resofro.2025.5\(1\)611](https://doi.org/10.59814/resofro.2025.5(1)611)
- Galán-Íñigo, A., Ruiz-Lázaro, J. y Jiménez-García, E. (2025). La inteligencia artificial en los procesos de evaluación en educación superior: un análisis bibliométrico (2014-2024). *Bordón, Revista de Pedagogía*, 77(3), 131-154. <https://doi.org/10.13042/Bordon.2025.107797>

- Hernández-León, N. y Rodríguez-Conde, M. J. (2024). Inteligencia artificial aplicada a la educación y la evaluación educativa en la universidad: introducción de sistemas de tutorización inteligentes, sistemas de reconocimiento y otras tendencias futuras. *Revista de Educación a Distancia*, 24(78). <http://dx.doi.org/10.6018/red.594651>
- Kiling, S. (2024). Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration. *Journal of Educational Technology & Online Learning*, 7(4), 521-540. <http://doi.org/10.31681/jetol.1492695>
- Llanos Torrico, B. A. (2025). La evaluación educativa en la era de la inteligencia artificial. *Revista Educación Superior*, 12(1), 73-88. <https://doi.org/10.53287/undf7848pz65o>
- López Escobar, C. A., Hernández Salas, C., Hernández Castillo, J. B., Cabrera Gómez, T. del C. y Maldonado Castillo, B. L. (2024). Evaluación de la eficiencia de las herramientas de evaluación académica: una revisión sistemática. *Ciencia Latina Revista Científica Multidisciplinar*, 8(4), 6848-6862. https://doi.org/10.37811/cl_rcm.v8i4.12868
- Miller, W. (2024). Adapting to AI: Reimagining the role of assessment professionals. *Intersection: A Journal at the Intersection of Assessment and Learning*, 5(4), 99-113. <https://tinyurl.com/24c4678a>
- Mpolomoka, D. L. (2025). Utilizing artificial intelligence for assessment in higher education. *Pedagogical Research*, 10(3), em0243. <https://doi.org/10.29333/pr/16677>
- Perplexity AI (s. f.). *Perplexity* [plataforma de preguntas y respuestas con IA]. <https://www.perplexity.ai>
- QuillBot (s. f.). *QuillBot* [herramienta de redacción y parafraseo]. <https://www.quillbot.com>