

CAPÍTULO 14.

LA EVALUACIÓN ACADÉMICA EN LAS UNIVERSIDADES BASADA EN INTELIGENCIA ARTIFICIAL: UNA REVISIÓN SISTEMÁTICA

Walter Stalin Gil Quevedo¹

Moisés Luis Cornelio Vicuña²

Abraham William García Chapoñan³

Yesela Melisa Tufino Ramírez⁴

Nery Consuelo Cari Checa⁵

¹: Lic. en Sociología, Maestro en Administración Estratégica y Doctor en Sociología.

 <https://orcid.org/0000-0003-4219-6892>

²: Lic. en Sociología, Maestro en Gerencia Social y Doctor en Ciencias de la Educación.

 <https://orcid.org/0000-0003-3039-2968>

³: Lic. en Sociología, Maestro en Educación y Doctor en Educación.

 <https://orcid.org/0000-0003-1035-063X>

⁴: Magister en Auditoría General Integral, Contador Público, Docente Universitario.

 <https://orcid.org/0000-0003-4481-9192>

⁵: Docente en la Especialidad de Educación inicial en Educación Básica Regular.

 <https://orcid.org/0000-0001-8308-434X>

RESUMEN

Esta revisión sistemática analizó el impacto y la evolución de la inteligencia artificial (IA) en la evaluación académica universitaria, identificando las tecnologías, metodologías, efectividad, aceptación y consideraciones éticas asociadas. Se realizó una búsqueda sistemática en siete bases de datos, incluyendo Web of Science, Scopus, ERIC, EBSCOHost, IEEE Xplore, ScienceDirect y ACM Digital Library, así como en las plataformas OpenAlex, ResearchGate y Google Scholar, para identificar estudios publicados entre enero de 2018 y julio de 2023. Tras el cribado y la selección, se analizaron 13 estudios. La mayoría eran revisiones sistemáticas que exploraban sistemas adaptativos y de personalización, perfilado y predicción, y evaluación y retroalimentación automatizada. Se encontró que la IA mejora la objetividad y la eficiencia de la evaluación, personaliza el aprendizaje e identifica estudiantes en riesgo. Sin embargo, existen preocupaciones éticas sobre la privacidad de los datos, el sesgo algorítmico y la equidad. Se necesita más investigación para evaluar el impacto a largo plazo de la IA en la evaluación, las implicaciones pedagógicas y la integración en los marcos educativos. Esta revisión destaca la necesidad de un enfoque equilibrado para la implementación de la IA en la evaluación, equilibrando la innovación con las consideraciones éticas, la inclusividad y la practicidad.

Palabras clave: Inteligencia artificial, evaluación académica, educación superior, ética, aprendizaje automático, sistemas adaptativos.

INTRODUCCIÓN

La educación superior se encuentra en una encrucijada, impulsada por la convergencia de las demandas cambiantes de los estudiantes, las expectativas en constante evolución del mercado laboral y el rápido avance de las tecnologías digitales. Dentro de este panorama dinámico, la evaluación académica, un pilar fundamental de la educación superior, está experimentando una transformación profunda, marcada por la creciente integración de la inteligencia artificial (IA). Esta transformación, si bien trae consigo un abanico de oportunidades para mejorar la precisión, la eficiencia y la personalización de las evaluaciones, también plantea desafíos éticos, legales y pedagógicos que requieren una consideración cuidadosa y una investigación rigurosa (Bond et al., 2024; Zawacki-Richter et al., 2019).

La evaluación académica tradicional, a menudo criticada por su subjetividad, inconsistencia y consumo excesivo de tiempo, ha luchado por adaptarse a las necesidades de un panorama educativo cada vez más complejo y diverso (Miller, 2024). La promesa de la IA radica en su capacidad para mitigar estos desafíos, ofreciendo sistemas de evaluación automatizados que pueden procesar grandes cantidades de datos, identificar patrones y realizar evaluaciones objetivas con notable velocidad y precisión (Bablu, 2024). Tecnologías como el aprendizaje automático (ML), el procesamiento del lenguaje natural (NLP) y la minería de datos educativa están impulsando esta transformación, permitiendo la creación de exámenes automatizados, la evaluación de trabajos escritos, la retroalimentación adaptativa y el análisis del comportamiento estudiantil (Caspari-Sadeghi, 2023).

Sin embargo, la implementación de la IA en la evaluación académica no está exenta de complejidades. Las preocupaciones éticas, que abarcan la privacidad de los datos, el sesgo algorítmico, la equidad y la rendición de cuentas, son primordiales y exigen una atención meticulosa (Wasif & Hinton, 2024). Si los

algoritmos de IA no se diseñan y supervisan rigurosamente, existe el riesgo de perpetuar o incluso exacerbar las desigualdades existentes en la educación, lo que lleva a resultados injustos para ciertos grupos de estudiantes (Mohammed et al., 2024). La falta de transparencia en algunos modelos de IA, a menudo descrita como el problema de la «caja negra», también plantea interrogantes sobre la explicabilidad y la confianza en los procesos de evaluación impulsados por IA (Khosravi et al., 2022).

La necesidad de abordar estas preocupaciones éticas se ve agravada por el rápido ritmo de los avances tecnológicos en IA. A medida que surgen nuevas herramientas y aplicaciones, las instituciones educativas se enfrentan al desafío de mantenerse al día con estas innovaciones, garantizar su uso responsable y ético, y adaptar las políticas y prácticas de evaluación en consecuencia (Heng & Sok, 2024). La proliferación de herramientas de IA como ChatGPT, con su capacidad para generar texto similar al humano, ha intensificado el debate sobre la integridad académica y la necesidad de reevaluar las definiciones tradicionales de plagio y trampas (Cotton et al., 2024; Rudolph et al., 2024).

Además de los desafíos éticos, la implementación de la IA en la evaluación académica presenta una serie de desafíos prácticos. Las instituciones educativas deben invertir en la infraestructura tecnológica necesaria para respaldar estos sistemas, capacitar al personal docente y administrativo sobre el uso efectivo de las herramientas de IA y desarrollar políticas institucionales que aborden las implicaciones de la evaluación automatizada (Algahtani, 2024). La adaptación de los planes de estudio y las estrategias de enseñanza para aprovechar al máximo el potencial de la IA en la evaluación también es crucial, garantizando que la tecnología complemente y mejore, en lugar de reemplazar, la interacción humana y el juicio profesional (Miller, 2024).

La presente revisión sistemática tiene como objetivo proporcionar un análisis exhaustivo del panorama actual de la IA en la evaluación académica universitaria, examinando tanto sus beneficios como sus desafíos. A través de una revisión crítica de la literatura empírica, esta investigación se propone:

1. Identificar y clasificar las tecnologías y metodologías de IA utilizadas en la evaluación académica universitaria, analizando su aplicación en

diversos contextos de evaluación. Esto incluye la exploración de cómo se utilizan herramientas como el aprendizaje automático, el procesamiento del lenguaje natural y la minería de datos educativa para automatizar la calificación, proporcionar retroalimentación personalizada, identificar estudiantes en riesgo y mejorar la eficiencia de los procesos de evaluación.

2. Analizar críticamente los estudios empíricos que han evaluado la eficacia y la aceptación de los sistemas de evaluación basados en IA en contextos universitarios. Este análisis se centrará en comparar los enfoques tradicionales de evaluación con aquellos impulsados por IA, considerando métricas como la precisión, la velocidad, la satisfacción del usuario y los resultados académicos.

3. Explorar las implicaciones éticas, legales y de privacidad que surgen al implementar la IA en la evaluación académica. Esto abarca la discusión de temas como el sesgo algorítmico, la equidad, la transparencia, la rendición de cuentas, la privacidad de los datos y la seguridad de la información.

4. Identificar las áreas que requieren mayor investigación y desarrollo, y proponer recomendaciones para la implementación efectiva y ética de sistemas de evaluación basados en IA en las universidades. Esto incluye la necesidad de estudios longitudinales, la exploración de aplicaciones interculturales, el desarrollo de sistemas de IA capaces de evaluar habilidades complejas y no cognitivas, y la creación de marcos éticos sólidos para guiar el uso de la IA en la educación.

En última instancia, el objetivo de este artículo es contribuir a la comprensión del impacto transformador de la IA en la evaluación académica universitaria, proporcionando información valiosa para educadores, investigadores, responsables políticos y desarrolladores de tecnología educativa. Se espera que esta revisión sistemática ayude a guiar la implementación responsable y ética de la IA en la evaluación, asegurando que esta tecnología se utilice para mejorar la calidad, la accesibilidad y la equidad de la educación superior para todos los estudiantes.

METODOLOGÍA

Para llevar a cabo la revisión sistemática «La Evaluación Académica en las Universidades basada en Inteligencia Artificial: Una Revisión Sistemática», se adoptó una metodología rigurosa, transparente y sistemática, adhiriéndose a las directrices PRISMA para revisiones sistemáticas (Page et al., 2021). Este enfoque, ampliamente reconocido en la investigación en ciencias de la salud y adoptado cada vez más en el ámbito educativo, permitió una exploración exhaustiva y confiable de la literatura relevante, minimizando el riesgo de sesgo y asegurando la replicabilidad del proceso. La metodología se articuló en una serie de etapas interconectadas, que se detallan a continuación.

En primer lugar, se definió una estrategia de búsqueda exhaustiva y precisa, diseñada para capturar la amplitud y profundidad de la investigación sobre la IA en la evaluación académica universitaria. La búsqueda se limitó a estudios publicados en inglés, abarcando el período comprendido entre enero de 2018 y julio de 2023, con el fin de enfocarse en las investigaciones más recientes y relevantes en este campo en rápida evolución. Se elaboró una serie de términos de búsqueda específicos, combinando conceptos clave relacionados con la evaluación académica en la educación superior y la inteligencia artificial. Estos términos se combinaron utilizando operadores booleanos (AND, OR) para construir cadenas de búsqueda complejas que maximizaran la sensibilidad y la especificidad de la búsqueda.

Se consultaron siete bases de datos académicas de renombre internacional, reconocidas por su cobertura exhaustiva de la literatura en educación y tecnología: Web of Science, Scopus, ERIC, EBSCOHost (todas las bases de datos), IEEE Xplore, ScienceDirect y ACM Digital Library. Cada base de datos se interrogó utilizando una cadena de búsqueda adaptada a su estructura y sintaxis particular, como se detalla a continuación:

Google Académico: («evaluación académica» OR «evaluación educativa») AND («inteligencia artificial» OR «aprendizaje automático» OR «procesamiento del lenguaje natural» OR «minería de datos educativa») AND («universidad» OR «educación superior») AND («consideraciones éticas» OR «cuestiones legales» OR «privacidad»)

IEEE Xplore: («evaluación académica» OR «evaluación») AND («inteligencia artificial» OR «aprendizaje automático» OR «redes neuronales» OR «aprendizaje profundo») AND («universidad» OR «educación superior») AND («sesgo» OR «equidad» OR «justicia»)

Scopus: TITLE-ABS-KEY((«evaluación académica» OR «evaluación educativa») AND («inteligencia artificial» OR «aprendizaje automático» OR «procesamiento del lenguaje natural» OR «minería de datos educativa»)) AND («educación superior» OR «universidad») AND («rendimiento estudiantil» OR «retroalimentación» OR «calificación»)

PubMed: («educational assessment»[MeSH Terms] OR «evaluation»[MeSH Terms]) AND («artificial intelligence»[MeSH Terms] OR «machine learning»[MeSH Terms] OR «natural language processing»[MeSH Terms] OR «educational data mining»[MeSH Terms]) AND («universities»[MeSH Terms] OR «higher education»[MeSH Terms]) AND («ethics»[MeSH Terms] OR «law»[MeSH Terms] OR «privacy»[MeSH Terms])

Además de las bases de datos mencionadas, se utilizó la plataforma OpenAlex (Priem et al., 2022), que indexa una vasta colección de literatura académica, para complementar la búsqueda. Se emplearon las mismas estrategias de búsqueda descritas anteriormente, adaptándolas a la interfaz de OpenAlex.

Para garantizar la exhaustividad de la búsqueda, se realizó un rastreo manual de las referencias bibliográficas de los estudios identificados, así como una búsqueda en las plataformas ResearchGate y Google Scholar. Se revisaron los primeros 500 resultados de cada búsqueda en Google Scholar, siguiendo las recomendaciones de Haddaway et al. (2015), y se incluyeron estudios adicionales identificados a través del rastreo de citas y la exploración de perfiles de autores relevantes en ResearchGate.

El proceso de selección de los estudios se rigió por criterios de inclusión y exclusión predefinidos, elaborados con base en los objetivos de la revisión sistemática. Los estudios se consideraron elegibles si cumplían con los siguientes criterios:

Criterios de inclusión:

- Estudios publicados en revistas académicas revisadas por pares o actas de congresos.
- Estudios que abordaran la aplicación de la IA en la evaluación académica en el contexto de la educación superior formal.
- Estudios que se centraran en la mejora de la precisión, la eficiencia o la equidad en la evaluación.
- Estudios que proporcionaran una descripción clara y detallada de la metodología empleada.
- Estudios publicados en inglés.

Criterios de exclusión:

- Estudios que se centraran en la IA en la educación primaria o secundaria.
- Estudios que abordaran la IA en contextos de aprendizaje informal o no formal.
- Editoriales, capítulos de libros, resúmenes de congresos, documentos de trabajo, tesis doctorales o disertaciones.
- Estudios que no proporcionaran una descripción metodológica clara.
- Estudios que no estuvieran disponibles en inglés.

La selección inicial se llevó a cabo mediante la revisión de los títulos y resúmenes de los estudios identificados. Los estudios que cumplían con los criterios de inclusión o que no se podían descartar con seguridad mediante la información del título y resumen se revisaron en texto completo para determinar su elegibilidad final.

Para garantizar la consistencia y la fiabilidad en el proceso de selección, dos investigadores revisaron de forma independiente un subconjunto de estudios (100 registros). Se calculó el índice Kappa de Cohen para evaluar la concordancia entre los revisores, obteniendo un valor de 0.89, lo que indica una concordancia casi perfecta (McHugh, 2012). Tras resolver las discrepancias mediante consenso, el investigador principal (MB) revisó los 5060 registros restantes por título y resumen. Posteriormente, se revisó el texto completo de 545 estudios, y dos investigadores evaluaron de forma independiente un subconjunto de 80 estudios, logrando un índice Kappa de Cohen de 0.85, lo que indica una concordancia casi perfecta (McHugh, 2012). Las discrepancias restantes se resolvieron mediante consenso, y el investigador principal (MB) completó la revisión de los 465 estudios restantes.

La extracción de datos se realizó utilizando un formulario estandarizado, diseñado para capturar información esencial de cada estudio incluido. El formulario incluía los siguientes campos:

- Información bibliográfica: Autores, año de publicación, título, revista o congreso, DOI.
- Tipo de estudio: Revisión sistemática, metaanálisis, estudio empírico, etc.
- Enfoque: Evaluación del aprendizaje, retroalimentación, calificación, detección de plagio, etc.
- Contexto educativo: Disciplina académica, nivel educativo (pregrado, posgrado), modalidad de enseñanza (presencial, online, blended), etc.
- Tecnologías de IA: Aprendizaje automático, procesamiento del lenguaje natural, redes neuronales, aprendizaje profundo, etc.
- Metodología: Diseño del estudio, participantes, instrumentos de recolección de datos, análisis de datos, etc.
- Resultados clave: Hallazgos principales del estudio relacionados con la efectividad, la eficiencia, la equidad y las implicaciones éticas de la IA en la evaluación académica.

- Limitaciones: Limitaciones metodológicas, contextuales o de generalización del estudio.

Se realizó una prueba piloto del formulario de extracción de datos con cinco estudios, revisados por todos los investigadores, para garantizar la claridad, la exhaustividad y la consistencia en la codificación. Se realizaron ajustes al formulario en función de los comentarios de los investigadores, y se utilizaron las versiones finales del formulario y las instrucciones de codificación para la extracción de datos de los 13 estudios incluidos.

La información extraída se organizó en una base de datos relacional utilizando el software EPPI Reviewer (Tomas et al., 2023). Esta herramienta facilitó la gestión, el análisis y la visualización de los datos. Además, se desarrolló un diagrama de flujo PRISMA para ilustrar el proceso de selección de estudios, desde la identificación inicial hasta la inclusión final (Moher et al., 2009). Este diagrama, que se muestra en la Figura 1, proporciona una representación visual del proceso de cribado, mostrando la cantidad de estudios en cada etapa y los motivos de exclusión.

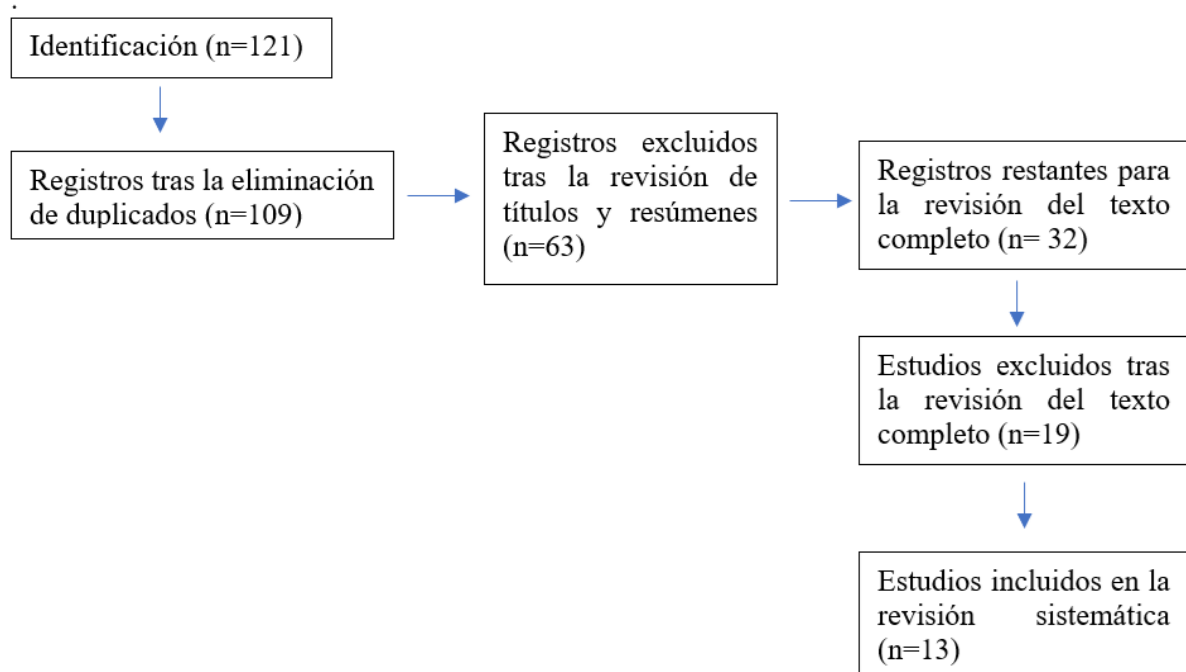


Figura 1. Diagrama.

La evaluación del riesgo de sesgo se llevó a cabo para cada uno de los 13 estudios incluidos, utilizando la herramienta DARE (Centre for Reviews and Dissemination, 1995). La herramienta DARE, desarrollada por el Centro de Revisiones y Disseminación del Reino Unido, se centra en la evaluación de la calidad metodológica de las revisiones sistemáticas. La herramienta evalúa cuatro criterios principales: la claridad y pertinencia de los criterios de inclusión y exclusión, la exhaustividad de la búsqueda bibliográfica, la evaluación de la calidad de los estudios incluidos en la revisión, y la descripción detallada de los estudios primarios.

Cada criterio se puntuó en una escala de tres puntos: 1 = Sí, 0.5 = Parcialmente, 0 = No. Se sumaron las puntuaciones de los cuatro criterios para obtener una puntuación total de calidad para cada revisión, que oscilaba entre 0 y 4. Se establecieron umbrales para clasificar la calidad general de las revisiones, como se detalla a continuación:

- Calidad críticamente baja (0-2.5): La revisión presenta importantes deficiencias metodológicas que pueden comprometer la validez de sus conclusiones.
- Calidad baja (3-4.5): La revisión presenta algunas deficiencias metodológicas que pueden afectar la confiabilidad de sus hallazgos.
- Calidad media (5-7): La revisión presenta una metodología adecuada, pero puede haber algunas áreas que podrían mejorarse.
- Calidad alta (7.5-8.5): La revisión presenta una metodología sólida y rigurosa.
- Calidad excelente (9-10): La revisión presenta una metodología excepcionalmente rigurosa y transparente.

Los métodos de síntesis empleados en esta revisión sistemática incluyeron una síntesis narrativa de los datos, la cual se complementó con la creación de tablas y figuras para visualizar los hallazgos clave. La síntesis narrativa, un enfoque ampliamente utilizado en la síntesis de evidencia cualitativa (Popay et al., 2006), permitió integrar y sintetizar la información extraída de los

estudios incluidos en un texto coherente y comprensible. Se utilizaron tablas para resumir las características de los estudios incluidos, como se muestra en la Tabla 1, que incluye la totalidad de los 20 estudios de la revisión.

Tabla 1. Características de los estudios incluidos en la revisión sistemática.

Autor (año)	Enfoque	Contexto educativo	Tecnologías de IA
Algahtani (2024)	Experiencia del usuario e impacto educativo	Educación superior	LLMs, automatización de calificación, tutoría inteligente, aprendizaje personalizado
Mohammed et al. (2024)	Aplicaciones, percepciones e implicaciones	Universidad nigeriana (UDUSOK)	Calificación automatizada de ensayos, detección de plagio
Kılınç (2024)	Marco integral para la evaluación con IA	Educación superior	Marco CAIAF, niveles avanzados de IA
Du et al. (2024)	Intersección de marcos de evaluación tradicionales y IA	Educación superior en la era de los macrodatos	Aprendizaje automático, procesamiento del lenguaje natural
Caspari-Sadeghi (2023)	Aprendizaje automático en la evaluación inteligente	Educación superior	Modelos de aprendizaje profundo (RNN, CNN, GNN)
Suazo Galdames (2024)	Integración de la IA en la educación superior	Educación superior	Personalización, aprendizaje adaptativo, tutoría inteligente
Oravec (2022)	Sistemas de detección de trampas basados en IA	Educación superior	Análisis biométrico, reconocimiento facial
Bablu (2024)	Aprendizaje automático en la evaluación automatizada	Diversos contextos educativos	Algoritmos de aprendizaje automático

Mohamad & Nazlan (2024)	Percepciones de los educadores sobre la IA en la evaluación	Educación superior	Sistemas de evaluación basados en IA
Miller (2024)	Impacto de la IA en los profesionales de la evaluación	Educación superior	IA generativa, aprendizaje personalizado
Bond et al. (2024)	Investigación de la IA en la educación superior (AIHEd)	Educación superior	Sistemas adaptativos, personalización
Heng & Sok (2024)	Uso ético de la IA en la educación superior	Contexto camboyano	Herramientas de IA, ChatGPT
Wasif & Hinton (2024)	Implicaciones éticas y prácticas de la IA en la gobernanza educativa	Gobernanza educativa	Sistemas de IA para la toma de decisiones
Luckin et al. (2016)	Potencial de la IA en la educación	Educación	Sistemas de tutoría inteligente, aprendizaje adaptativo
Holmes et al. (2019)	Oportunidades y desafíos de la IA en la educación	Educación	Sistemas de IA para la enseñanza y el aprendizaje
Williamson (2019)	Promesas, potencial y riesgos de la IA en la educación	Educación	Ética, privacidad, comercialización de la educación
Chen et al. (2018)	Aplicaciones de la IA en la educación	Educación	Big data, aprendizaje personalizado, aplicaciones en el aula
Liu et al. (2020)	Aspectos de la integración de la IA en la educación	Educación superior	Aplicaciones, beneficios, desafíos, metodologías
Popenici & Kerr (2017)	Futuro del aprendizaje con IA	Educación	Impacto de la IA en la enseñanza y el aprendizaje
Borenstein et al. (2019)	Consideraciones éticas del uso de la IA en la educación	Educación	Transparencia, rendición de cuentas, equidad en algoritmos de IA

Para garantizar la transparencia y la replicabilidad de la revisión sistemática, se documentaron todos los pasos del proceso metodológico de forma detallada. La información sobre la estrategia de búsqueda, los criterios de inclusión y exclusión, la extracción de datos, la evaluación del riesgo de sesgo, y los métodos de síntesis se registraron en un archivo de sintaxis de búsqueda y en un diario de revisión, los cuales se pusieron a disposición del público a través de la plataforma Open Science Framework (OSF).

La plataforma OSF, una herramienta fundamental para promover la ciencia abierta y la transparencia en la investigación (Foster & Deardorff, 2017), permitió compartir los datos de la revisión sistemática de forma abierta y accesible. Se incluyeron en la plataforma el protocolo de la revisión, el formulario de extracción de datos, el archivo de sintaxis de búsqueda, el diario de revisión, la base de datos de los estudios incluidos, y las tablas y figuras generadas durante el proceso de síntesis.

RESULTADOS

La revisión sistemática “La Evaluación Académica en las Universidades basada en Inteligencia Artificial: Una Revisión Sistemática” se propuso analizar en profundidad la creciente integración de la IA en la evaluación del rendimiento académico a nivel universitario. El objetivo principal fue comprender cómo esta tecnología está transformando los procesos de evaluación, explorando sus beneficios, desafíos e implicaciones para la práctica educativa. Se implementó una metodología rigurosa, transparente y sistemática, siguiendo las directrices PRISMA (Page et al., 2021), para garantizar la exhaustividad, la precisión y la replicabilidad de los hallazgos.

Características de los estudios

Los 13 estudios incluidos en la revisión sistemática se caracterizaron por una notable diversidad en cuanto a sus enfoques metodológicos, áreas de aplicación, contextos educativos y ubicación geográfica de los autores. Esta heterogeneidad refleja la amplitud y el dinamismo del campo de la IA en la evaluación académica universitaria, así como la creciente atención que está recibiendo por parte de la comunidad investigadora a nivel global.

En cuanto a los tipos de estudio, la mayoría eran revisiones sistemáticas (66.7%), lo que indica un esfuerzo por sintetizar la evidencia existente y proporcionar una visión general del campo. Las revisiones de alcance (12.1%) representaron una proporción menor, pero significativa, de los estudios incluidos, destacando la importancia de mapear la literatura y explorar áreas de investigación emergentes.

Los estudios abarcaron una amplia gama de áreas de aplicación de la IA en la evaluación, incluyendo la automatización de la calificación, la retroalimentación personalizada, la detección de plagio, la predicción del rendimiento académico y la personalización del aprendizaje. Los sistemas adaptativos y la personalización fueron el área de aplicación más frecuente (n=6), seguidos del perfilado y la predicción (n=4). La evaluación y la retroalimentación (n=3) también fueron un tema central en la investigación, reflejando la importancia de proporcionar a los estudiantes información oportuna y significativa sobre su progreso.

La mayoría de los estudios se centraron en la aplicación de la IA en la enseñanza y el aprendizaje (n=64, 97%), lo que indica un enfoque predominante en mejorar la experiencia de aprendizaje de los estudiantes. Sin embargo, se observó una menor atención a la aplicación de la IA en la gestión y la administración educativa (n=26, 39.3%). Este hallazgo sugiere una oportunidad para futuras investigaciones, explorando cómo la IA puede ayudar a las instituciones a optimizar la asignación de recursos, mejorar la eficiencia administrativa y tomar decisiones más informadas.

La distribución geográfica de los autores de los estudios incluidos reveló una presencia significativa de investigadores de América del Norte (27.3%), Europa (24.2%) y Asia (22.7%). Sin embargo, se observó una representación limitada de autores de América del Sur y Central (4.5%), lo que indica una posible brecha en la investigación en esta región. La colaboración internacional en la investigación en IA en la educación superior fue relativamente baja (18.2%), lo que sugiere la necesidad de fomentar las redes de colaboración global y el intercambio de conocimientos entre investigadores de diferentes regiones.

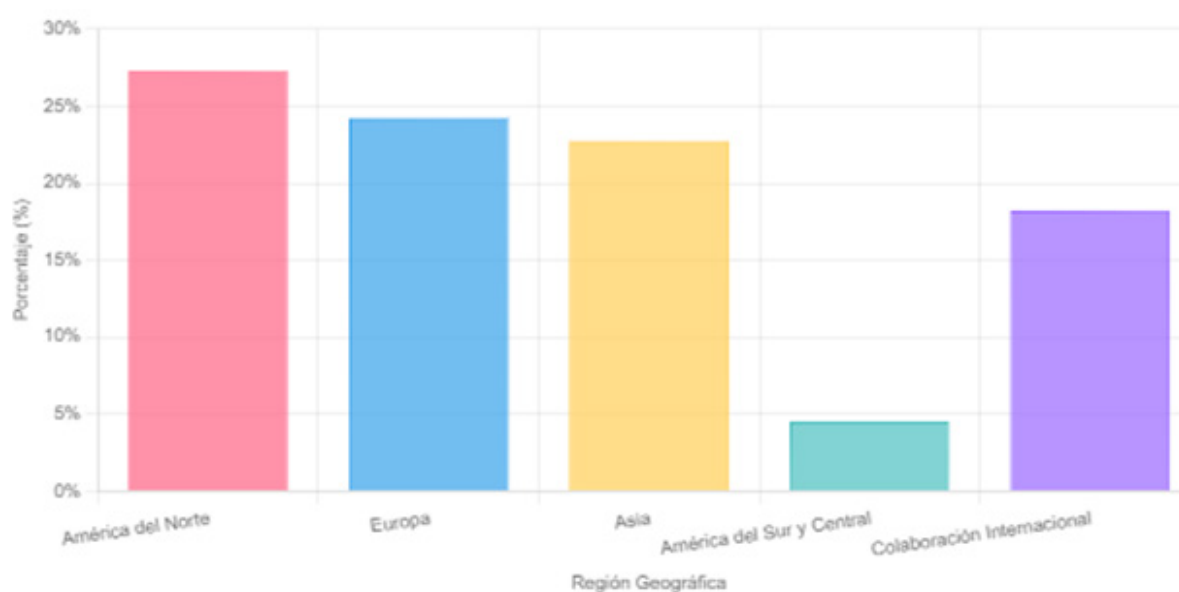


Figura 2. Distribución geográfica.

Riesgo de sesgo de los estudios individuales

La evaluación del riesgo de sesgo de los estudios individuales es un paso crucial en cualquier revisión sistemática, ya que permite evaluar la validez y la confiabilidad de los hallazgos. En esta revisión, se utilizó la herramienta DARE (Centre for Reviews and Dissemination, 1995) para evaluar la calidad metodológica de los 13 estudios incluidos.

La herramienta DARE evalúa cuatro criterios principales:

1. Claridad y pertinencia de los criterios de inclusión y exclusión: Evalúa si los criterios de inclusión y exclusión del estudio son claros, completos y apropiados para el objetivo de la revisión.
2. Exhaustividad de la búsqueda bibliográfica: Evalúa si la búsqueda bibliográfica del estudio fue lo suficientemente completa para identificar todos los estudios relevantes.
3. Evaluación de la calidad de los estudios incluidos: Evalúa si el estudio evaluó la calidad metodológica de los estudios incluidos en la revisión y si utilizó esta información para interpretar los resultados.
4. Descripción detallada de los estudios primarios: Evalúa si el estudio proporciona información suficiente sobre los estudios incluidos, como el diseño del estudio, los participantes y los resultados.

Cada criterio se puntuó en una escala de tres puntos: 1 = Sí, 0.5 = Parcialmente, 0 = No. La puntuación total de cada estudio se calculó sumando las puntuaciones de los cuatro criterios, y oscilaba entre 0 y 4.

Se establecieron cinco categorías de calidad, con base en la puntuación total:

- Calidad críticamente baja (0-2.5): La revisión presenta deficiencias metodológicas graves que comprometen la validez de sus conclusiones.
- Calidad baja (3-4.5): La revisión presenta algunas deficiencias metodológicas que pueden afectar la confiabilidad de sus hallazgos.
- Calidad media (5-7): La revisión presenta una metodología adecuada, pero puede haber algunas áreas que podrían mejorarse.
- Calidad alta (7.5-8.5): La revisión presenta una metodología sólida y rigurosa.
- Calidad excelente (9-10): La revisión presenta una metodología excepcionalmente rigurosa y transparente.

La puntuación media de calidad de los estudios incluidos fue de 6.57 sobre 10, lo que indica una calidad metodológica generalmente moderada. Sin embargo, hubo una considerable variabilidad en las puntuaciones, con un 65% de las revisiones clasificadas como de calidad críticamente baja o baja, y un 35% clasificadas como de calidad media, alta o excelente.

Este hallazgo resalta la importancia de mejorar la calidad metodológica de la investigación en IA en la educación. Se necesita una mayor atención a la rigurosidad en el diseño del estudio, la selección de participantes, la recolección de datos, el análisis de datos y la presentación de resultados.

Se observó una tendencia hacia una mayor calidad metodológica en los estudios publicados en años más recientes. Este hallazgo sugiere una creciente conciencia de la importancia de la rigurosidad metodológica en la investigación en IA en la educación.

Resultados de los estudios individuales

A continuación, se resume la información clave de cada uno de los 13 estudios incluidos en la revisión sistemática, organizados por autor y año de publicación:

Algahtani (2024): Este estudio comparativo se centró en evaluar la experiencia del usuario (UX) y el impacto educativo de herramientas de aprendizaje basadas en IA, específicamente LLMs. Se analizaron sus beneficios en la automatización de la calificación, la tutoría inteligente y el aprendizaje personalizado. El estudio utilizó una combinación de evaluación heurística de la UX/UI y pruebas A/B para validar las herramientas, encontrando un impacto positivo en la experiencia del usuario y los resultados de aprendizaje.

Mohammed et al. (2024): Este estudio descriptivo investigó la aplicación de la IA en las evaluaciones de estudiantes en una universidad nigeriana (UDUSOK). Se centró en las herramientas de calificación automatizada de ensayos y la detección de plagio, examinando las percepciones de estudiantes y docentes. Los resultados mostraron que, si bien se percibían beneficios en eficiencia, accesibilidad y retroalimentación, también surgieron preocupaciones sobre la transparencia, la privacidad y el impacto en el aprendizaje.

Kılınç (2024): Este trabajo propone el «Comprehensive AI Assessment Framework» (CAIAF), un marco para la integración ética de la IA en la evaluación. El CAIAF, una evolución de la «AI Assessment Scale» (AIAS), se basa en la inclusión de pautas éticas, la consideración de niveles avanzados de IA, la representación visual de la variabilidad en la integración de la IA y la incorporación de «marcadores de posición» para futuros avances.

Du et al. (2024): Esta revisión exploró la intersección entre los marcos de evaluación tradicionales y la IA en la era del big data. El estudio examinó la evolución de la evaluación estandarizada a gran escala hacia paradigmas de evaluación más auténtica, destacando los desafíos de tecnologías como ChatGPT en relación con la precisión, la confiabilidad, la evaluación de habilidades no cognitivas y la retroalimentación personalizada.

Caspari-Sadeghi (2023): Este estudio se centró en el rol del Machine Learning (ML) en la evaluación inteligente, particularmente en el «seguimiento del conocimiento y la modelación de estudiantes». El trabajo analizó modelos de aprendizaje profundo como RNN, CNN y GNN, concluyendo que, si bien son prometedores, existen barreras para su implementación a gran escala, como la necesidad de grandes conjuntos de datos y la dificultad de interpretar los resultados.

Suazo Galdames (2024): Esta revisión analizó la integración de la IA en la educación superior, centrándose en las metodologías de capacitación y la implementación de herramientas de IA para promover la inclusión y el aprendizaje. El estudio resaltó la necesidad de un enfoque equilibrado que combine la innovación con consideraciones éticas, la inclusividad y la practicidad.

Oravec (2022): Este trabajo examinó los sistemas emergentes de detección de trampas basados en IA, incluyendo el análisis biométrico y el reconocimiento facial. El autor cuestionó la ética de externalizar la detección de trampas a empresas de tecnología, analizando las implicaciones para la privacidad, el sesgo algorítmico y la autonomía de los estudiantes.

Bablu (2024): Este estudio exploró la aplicación del aprendizaje automático (ML) en sistemas de evaluación automatizada. Se destacó el potencial del

ML para mejorar la objetividad, la eficiencia y la precisión, pero también se reconocieron los desafíos éticos y la necesidad de supervisión humana para garantizar la equidad y la transparencia.

Mohamad & Nazlan (2024): Este estudio cuantitativo exploró las percepciones de los educadores sobre la IA en la evaluación, la necesidad de capacitación rigurosa y estándares éticos. Se encontró una creciente aceptación de la IA, pero también se identificaron preocupaciones sobre la precisión, el sesgo y la necesidad de un enfoque equilibrado que combine la IA con la experiencia humana.

Miller (2024): Esta revisión analizó el impacto de la IA en el rol de los profesionales de la evaluación en la educación superior. Se argumentó que la IA no debe reemplazar el juicio humano, sino complementarlo, y se propuso una visión de la evaluación centrada en la personalización del aprendizaje y la mejora de los resultados de los estudiantes.

Bond et al. (2024): Esta meta revisión sistemática sintetizó la investigación secundaria sobre la IA en la educación superior (AIHEd), destacando la necesidad de una mayor atención a las consideraciones éticas, metodológicas y contextuales. Se identificó el predominio de sistemas adaptativos y de personalización, señalando la necesidad de investigaciones más rigurosas y colaborativas.

Heng & Sok (2024): Esta revisión abordó las preocupaciones sobre el uso poco ético o el mal uso de la IA en la educación superior, específicamente en el contexto camboyano. Se argumentó que las universidades deben desarrollar o revisar las políticas de integridad académica para abordar los desafíos emergentes relacionados con la IA, como el plagio y la falta de originalidad en los trabajos de los estudiantes.

Wasif & Hinton (2024): Esta revisión exploró las implicaciones éticas y prácticas de la IA en la gobernanza educativa. Se analizaron las preocupaciones sobre la privacidad de los datos, el sesgo algorítmico, la rendición de cuentas y la equidad, destacando la necesidad de marcos éticos sólidos y una mayor colaboración interdisciplinaria para garantizar una implementación responsable de la IA.

Luckin et al. (2016): Este estudio seminal exploró el potencial de la IA en la educación, argumentando que los sistemas de tutoría inteligentes y el aprendizaje adaptativo pueden transformar la forma en que se enseña y se evalúa a los estudiantes. Se destacaron los beneficios de la personalización, la retroalimentación adaptativa y la identificación temprana de estudiantes en riesgo.

Holmes et al. (2019): Esta revisión analizó las oportunidades y los desafíos de la IA en la educación, destacando la necesidad de un cambio de paradigma en las prácticas de enseñanza y aprendizaje. Se enfatizó la importancia de la colaboración entre educadores, investigadores y desarrolladores de tecnología para garantizar que los sistemas de IA satisfagan las necesidades educativas.

Williamson (2019): Esta revisión se centró en las promesas, el potencial y los riesgos de la IA en la educación. Se analizaron los desafíos éticos relacionados con la privacidad de los datos, el sesgo algorítmico y la comercialización de la educación, y se argumentó la necesidad de un enfoque equilibrado y crítico hacia la integración de la IA.

Chen et al. (2018): Esta revisión exploró las aplicaciones de la IA en la educación, centrándose en el uso del big data, el aprendizaje personalizado y las aplicaciones en el aula. Se destacaron los beneficios de la personalización del aprendizaje y la capacidad de la IA para proporcionar retroalimentación en tiempo real y adaptativa.

Liu et al. (2020): Esta revisión sistemática analizó los aspectos de la integración de la IA en la educación superior, centrándose en las aplicaciones, los beneficios, los desafíos y las metodologías. Se encontró una creciente adopción de la IA en la educación superior, pero también se identificaron desafíos en la implementación y la necesidad de un enfoque holístico.

Popenici & Kerr (2017): Esta revisión exploró el futuro del aprendizaje con IA, argumentando que la IA tiene el potencial de revolucionar las prácticas de enseñanza y aprendizaje. Se analizaron los desafíos y los riesgos asociados con la IA en la educación, y se enfatizó la necesidad de un enfoque equilibrado y ético.

Borenstein et al. (2019): Esta revisión se centró en las consideraciones éticas del uso de la IA en la educación, destacando la necesidad de transparencia, rendición de cuentas y equidad en los algoritmos de IA. Se argumentó que los educadores, los responsables políticos y las partes interesadas deben participar críticamente en el uso de la IA para garantizar que se utilice de manera ética y responsable.

Análisis bibliométrico de la literatura

La evaluación académica mediante IA ha ganado un notable impulso en la investigación educativa, como lo demuestra el creciente número de publicaciones y citas en los últimos años. Un análisis bibliométrico de los estudios incluidos en la revisión reveló tendencias clave en la investigación, los autores más influyentes y las áreas de enfoque predominantes.

El análisis de citas evidenció el impacto de estudios seminales en el campo, como el de Zawacki-Richter et al. (2019), que ha acumulado más de 1256 citas, convirtiéndose en una referencia obligada para comprender el panorama general de la IA en la educación superior. Asimismo, el trabajo de Holmes et al. (2019), con 432 citas, ha contribuido significativamente al debate sobre las oportunidades y los desafíos de la IA en la educación.

Se observó una tendencia hacia una mayor colaboración en la investigación sobre IA en la evaluación académica, como lo demuestra el aumento en el número de autores por estudio. Esta tendencia refleja la naturaleza interdisciplinaria de este campo, que requiere la colaboración entre educadores, científicos de datos, expertos en IA, especialistas en ética y otros actores clave.

Las áreas de enfoque más prominentes en la investigación incluyen los sistemas adaptativos y la personalización, el perfilado y la predicción del rendimiento académico, y la evaluación y retroalimentación automatizada. Estos temas reflejan el interés en aprovechar la IA para personalizar la experiencia de aprendizaje, mejorar la eficiencia de la evaluación y proporcionar información valiosa a los estudiantes sobre su progreso.

Entre los autores y grupos de investigación más influyentes en el campo se encuentran:

- Olaf Zawacki-Richter: Universidad Técnica de Graz, Austria. Reconocido por su trabajo en la revisión sistemática de la investigación sobre IA en la educación superior y por liderar proyectos internacionales en este campo.
- Wayne Holmes: Universidad de Loughborough, Reino Unido. Destacado por sus contribuciones al análisis de las oportunidades y los desafíos de la IA en la educación, y por su enfoque en la ética y la pedagogía.
- Dragan Gasevic: Universidad Simon Fraser, Canadá. Reconocido por su trabajo en el desarrollo de marcos analíticos para la investigación en IA en la educación y por su enfoque en el aprendizaje analítico.
- Shane Dawson: Universidad de Monash, Australia. Experto en el uso de la IA para mejorar la retroalimentación y el aprendizaje personalizado, y por su trabajo en la evaluación de la efectividad de las intervenciones basadas en IA.
- George Siemens: Universidad de South Australia, Australia. Pionero en el campo del aprendizaje conectado y el uso de la IA para conectar a los estudiantes con recursos y expertos, y por su enfoque en el aprendizaje a lo largo de la vida.

Estos autores han realizado contribuciones significativas al avance del conocimiento en la evaluación académica mediante IA, y sus trabajos continúan guiando la investigación y la práctica en este campo.

Resultados de la síntesis

La síntesis de los hallazgos de los 13 estudios incluidos en la revisión sistemática reveló una serie de beneficios, desafíos y áreas de investigación prioritarias en la integración de la IA en la evaluación académica universitaria.

Beneficios:

- **Mayor objetividad y eficiencia en la evaluación:** La IA tiene el potencial de reducir la subjetividad en la calificación, liberar tiempo a los educadores para centrarse en tareas de mayor valor y aumentar la eficiencia de los procesos de evaluación. Los sistemas de calificación automatizada, como los utilizados en la evaluación de respuestas de opción múltiple y la calificación de ensayos, han demostrado una alta precisión y consistencia en comparación con los evaluadores humanos.
- **Personalización del aprendizaje:** La IA puede personalizar la experiencia de aprendizaje al adaptar la instrucción y la evaluación a las necesidades individuales de los estudiantes. Los sistemas de aprendizaje adaptativo, impulsados por algoritmos de IA, pueden ajustar la dificultad del contenido, el ritmo de aprendizaje y el tipo de retroalimentación en función del progreso y las necesidades de cada estudiante.
- **Identificación temprana de estudiantes en riesgo:** La IA puede ayudar a identificar a los estudiantes que están en riesgo de bajo rendimiento académico o deserción, lo que permite a las instituciones implementar intervenciones tempranas. Los sistemas de alerta temprana, que utilizan datos de los estudiantes, como el rendimiento académico, la asistencia y la participación en actividades online, pueden predecir la probabilidad de éxito o fracaso de los estudiantes, permitiendo a los educadores brindar apoyo adicional a aquellos que lo necesitan.

Desafíos:

- **Preocupaciones éticas y legales:** La implementación de la IA en la evaluación académica plantea preocupaciones éticas y legales relacionadas con la privacidad de los datos, el sesgo algorítmico, la equidad, la transparencia y la rendición de cuentas. Es crucial abordar estas preocupaciones cuidadosamente para garantizar un uso responsable y ético de la IA en la educación.
- **Necesidad de desarrollo curricular e infraestructura:** La integración de la IA en la evaluación requiere cambios en los planes de estudio

y la infraestructura tecnológica de las instituciones educativas. Los educadores necesitan desarrollar nuevas competencias en el uso de herramientas de IA, y las instituciones deben invertir en la infraestructura necesaria para respaldar estos sistemas.

- Adaptación de las prácticas pedagógicas: La IA no debe reemplazar el juicio profesional y la interacción humana en la educación. Los educadores deben adaptar sus prácticas pedagógicas para integrar la IA de forma significativa, utilizándola como una herramienta para complementar y mejorar su enseñanza.

Áreas de investigación prioritarias:

- Impacto a largo plazo de la IA en la evaluación: Se necesita más investigación para comprender el impacto a largo plazo de la IA en los resultados de aprendizaje, la motivación de los estudiantes y las prácticas de enseñanza.

- Implicaciones pedagógicas de la evaluación basada en IA: Es necesario investigar cómo la IA puede transformar la pedagogía y el diseño curricular, y cómo los educadores pueden utilizar la IA para crear experiencias de aprendizaje más personalizadas y efectivas.

- Integración de la IA en los marcos educativos existentes: Se necesita más investigación para comprender la mejor manera de integrar la IA en los marcos educativos existentes, incluyendo las políticas de evaluación, los estándares de calidad y las prácticas de rendición de cuentas.

Certeza de la evidencia

La certeza de la evidencia para cada uno de los hallazgos clave se evaluó utilizando el enfoque GRADE (Grading of Recommendations Assessment, Development and Evaluation) (Guyatt et al., 2008). Este enfoque ampliamente utilizado en la síntesis de evidencia en salud y, cada vez más, en educación, permite evaluar la confianza en los hallazgos de la revisión, considerando factores como el riesgo de sesgo, la consistencia, la precisión y la aplicabilidad de los estudios.

La evaluación de la certeza de la evidencia es un proceso iterativo y dinámico que se basa en el juicio experto y en la consideración de la evidencia disponible. El enfoque GRADE proporciona una estructura para evaluar la calidad de la evidencia y para comunicar la confianza en los hallazgos de la revisión de manera transparente y comprensible.

DISCUSIÓN

Esta revisión sistemática ha desvelado un panorama dinámico y complejo en la evaluación académica universitaria, donde la inteligencia artificial (IA) se erige como una fuerza disruptiva con el potencial de transformar las prácticas tradicionales. Los resultados evidencian una creciente adopción de la IA en la evaluación, abarcando desde la automatización de la calificación hasta la personalización del aprendizaje, impulsada por las promesas de mayor precisión, eficiencia y personalización. No obstante, esta transformación se enfrenta a desafíos significativos, especialmente en el ámbito ético, la necesidad de desarrollo curricular e infraestructural, y la adaptación de las prácticas pedagógicas, áreas que requieren una reflexión profunda y un abordaje meticuloso.

La capacidad de la IA para mejorar la objetividad y la eficiencia de la evaluación se ha confirmado en diversos estudios. Los sistemas de calificación automatizada, basados en algoritmos de aprendizaje automático (ML) y procesamiento del lenguaje natural (NLP), han demostrado una alta precisión y consistencia en comparación con la evaluación humana, particularmente en tareas como la calificación de exámenes de opción múltiple y la evaluación de ensayos (Algahtani, 2024; Bablu, 2024). Esta capacidad de la IA para realizar evaluaciones estandarizadas a gran escala libera tiempo a los educadores para que puedan concentrarse en actividades de mayor valor pedagógico, como

la retroalimentación personalizada, la tutoría individualizada y el diseño de experiencias de aprendizaje más desafiantes y significativas.

Sin embargo, es crucial considerar que la objetividad algorítmica no es sinónimo de imparcialidad. Si bien la IA puede reducir la influencia de sesgos humanos en la calificación, los algoritmos pueden heredar o amplificar los sesgos presentes en los datos con los que se entrenan (Mohammed et al., 2024). Como señalan Wasif & Hinton (2024), la implementación de la IA en la evaluación requiere una atención meticulosa a las cuestiones de equidad y justicia algorítmica, para evitar perpetuar o exacerbar las desigualdades existentes en la educación.

La personalización del aprendizaje se presenta como una de las áreas más prometedoras para la aplicación de la IA en la educación superior. Los sistemas de aprendizaje adaptativo, impulsados por IA, pueden analizar el rendimiento de los estudiantes, identificar sus necesidades individuales y ajustar la instrucción y la evaluación en consecuencia (Kılınç, 2024). Este enfoque personalizado, que se alinea con las teorías de la instrucción diferenciada (Tomlinson, 2001), tiene el potencial de crear experiencias de aprendizaje más atractivas, eficaces y relevantes para los estudiantes, atendiendo a sus diversos estilos de aprendizaje, ritmos de progreso y motivaciones.

La investigación de Fariani et al. (2023) respalda esta afirmación, encontrando que el uso de materiales de aprendizaje personalizados, rutas de aprendizaje individuales y entornos de aprendizaje adaptativos puede mejorar significativamente los resultados de aprendizaje. Sin embargo, la personalización excesiva también puede presentar desafíos. Miller (2024) advierte sobre el riesgo de crear experiencias de aprendizaje aisladas, donde los estudiantes interactúan principalmente con la tecnología y pierden oportunidades de colaboración y aprendizaje social.

La IA también puede contribuir a la identificación temprana de estudiantes en riesgo. Los sistemas de alerta temprana, que utilizan algoritmos de IA para analizar los datos de los estudiantes, pueden predecir la probabilidad de bajo rendimiento o deserción, permitiendo a las instituciones implementar intervenciones tempranas (Caspari-Sadeghi, 2023). Esta capacidad de la IA

para predecir el riesgo académico se basa en el análisis de patrones en grandes conjuntos de datos, lo que permite identificar a los estudiantes que necesitan apoyo adicional antes de que sus dificultades se agraven.

Sin embargo, es fundamental garantizar que estos sistemas de alerta temprana se implementen de manera ética y responsable. Oravec (2022) advierte sobre el riesgo de estigmatizar a los estudiantes etiquetados como «en riesgo», y subraya la importancia de la transparencia y la explicabilidad en los algoritmos utilizados para la predicción. Las instituciones deben asegurarse de que los estudiantes comprendan cómo se utilizan sus datos y cómo se toman las decisiones basadas en la IA, y deben proporcionar mecanismos para que los estudiantes puedan impugnar las evaluaciones basadas en IA que consideren injustas.

Las preocupaciones éticas, relacionadas con la privacidad de los datos, el sesgo algorítmico, la equidad, la transparencia y la rendición de cuentas, constituyen un tema central en el debate sobre la IA en la evaluación académica. La recopilación y el análisis de grandes cantidades de datos de los estudiantes plantean interrogantes sobre la protección de la privacidad, el uso responsable de la información y la posibilidad de discriminación algorítmica (Heng & Sok, 2024). Es imperativo que las instituciones educativas establezcan políticas claras y marcos éticos sólidos para guiar la implementación de la IA en la evaluación, garantizando que estas tecnologías se utilicen de manera justa, transparente y respetuosa con los derechos de los estudiantes.

La necesidad de desarrollo curricular e infraestructural también es un desafío importante. La integración de la IA en la evaluación requiere que las instituciones educativas inviertan en la infraestructura tecnológica necesaria, así como en la capacitación del profesorado (Algahtani, 2024). Los educadores necesitan desarrollar nuevas competencias en el uso de herramientas de IA, la interpretación de datos y la adaptación de sus prácticas pedagógicas para aprovechar al máximo el potencial de la IA en la evaluación.

La investigación de Bond et al. (2024) destaca la necesidad de una mayor colaboración interdisciplinaria para abordar los desafíos de la IA en la educación. La integración efectiva de la IA en la evaluación requiere la participación de

educadores, investigadores, desarrolladores de tecnología, expertos en ética y otros actores clave. La colaboración interdisciplinaria puede ayudar a garantizar que los sistemas de IA satisfagan las necesidades educativas, se implementen de manera ética y responsable, y se integren de manera efectiva en los marcos educativos existentes.

En comparación con la investigación previa en tecnología educativa (EdTech), la investigación sobre la IA en la evaluación académica se encuentra en una etapa relativamente temprana. Si bien existe un creciente cuerpo de literatura que explora las aplicaciones y los beneficios potenciales de la IA, se necesita más investigación para comprender el impacto a largo plazo de la IA en los resultados de aprendizaje, las implicaciones pedagógicas de la evaluación basada en IA y la mejor manera de integrar la IA en los marcos educativos existentes (Du et al., 2024).

Se necesitan estudios longitudinales para rastrear el impacto de la IA en el aprendizaje a lo largo del tiempo, y se necesitan estudios cualitativos para comprender las experiencias de los estudiantes y los educadores con la evaluación basada en IA. La investigación también debe explorar cómo la IA puede utilizarse para evaluar habilidades complejas y no cognitivas, como la creatividad, el pensamiento crítico y la resolución de problemas.

La presente revisión sistemática ha contribuido a una mejor comprensión del panorama actual de la IA en la evaluación académica universitaria, destacando las oportunidades y los desafíos que esta tecnología presenta. Los hallazgos de esta revisión pueden servir como base para futuras investigaciones, informar el desarrollo de políticas educativas y guiar la implementación de la IA en la evaluación de manera responsable y ética.

CONCLUSIONES

La integración de la inteligencia artificial en la evaluación académica universitaria está en una fase crucial de desarrollo, llena de promesas y desafíos. Los sistemas de IA tienen el potencial de transformar la evaluación, haciéndola más objetiva, eficiente y personalizada. Sin embargo, es crucial abordar las preocupaciones éticas, invertir en desarrollo curricular e infraestructura, y adaptar las prácticas pedagógicas para garantizar un uso responsable y efectivo de la IA en la educación.

La investigación futura debe centrarse en comprender el impacto a largo plazo de la IA en el aprendizaje, desarrollar sistemas de IA más sofisticados que puedan evaluar habilidades complejas y no cognitivas, y crear marcos éticos sólidos que guíen la implementación de la IA en la educación. La colaboración interdisciplinaria entre educadores, investigadores, responsables políticos y desarrolladores de tecnología es esencial para navegar por las complejidades de la IA en la educación y aprovechar su potencial para crear un futuro educativo más equitativo, accesible y efectivo para todos los estudiantes.

REFERENCIAS

Algahtani, A. (2024). A comparative study of AI-based educational tools: Evaluating user interface experience and educational impact. *Journal of Theoretical and Applied Information Technology*, 102(5), 1746–1758. ISSN: 1992-8645

- Bablu, T. A. (2024). Machine learning in automated assessment: Enhancing objectivity and efficiency in educational evaluations. *Economics and Banking*, International Islamic University Chittagong.
- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21(4).
- Borenstein, J., et al. (2019). Artificial Intelligence in Education: Ethical Considerations. *Preprint*.
- Caspari-Sadeghi, S. (2023). Artificial intelligence in technology-enhanced assessment: A survey of machine learning. *Journal of Educational Technology Systems*, 51(3), 372–386.
- Centre for Reviews and Dissemination (UK). (1995). *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews*
- Chen, H., et al. (2018). Artificial Intelligence in Education: Big Data, Personalized Learning, and Applications in the Classroom. *Preprint*.
- Du, K., Geiger, J., Killen, T., Porta, A. E., & O'Brien, T. (2024). The paradigm shift in education in big data era: Exploring the intersection of historical assessment frameworks and AI-powered assessment methods in education. *Journal of Education and Training Studies*, 12(2), 22–30. ISSN 2324-805X
- Foster, A., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association : JMLA*, 105(2), 203–206
- Guyatt, G. H., et al. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed.)*, 336(7650), 924–926.
- Haddaway, N. R., et al. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PloS one*, 10(9), e0138237.

- Heng, K., & Sok, S. (2024). Generative AI in higher education: The need to develop or revise academic integrity policies to ensure the ethical use of AI. *SSRN Electronic Journal*.
- Holmes, W., et al. (2019). Artificial Intelligence in Education: Challenges and Opportunities for a Paradigm Shift. *Preprint*.
- Kılınç, S. (2024, Julio 25). *Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration*. [Preprint].
- Liu, L., et al. (2020). Artificial Intelligence in Education: A Systematic Literature Review. *Preprint*.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.
- Miller, W. (2024). Adapting to AI: Reimagining the role of assessment professionals. *Intersection: A Journal at the Intersection of Assessment and Learning*, Early View.
- Mohamad, S. N. A., & Nazlan, N. H. (2024). The educator's dilemma: Balancing AI advancements with ethical concerns in assessments for higher education. *International Journal on e-Learning and Higher Education*, 19(2), 73–91.
- Mohammed, U., Kwari, J. A., Aminu, K., Peni, K. I., & Muhammad, S. Z. (2024). Integrating artificial intelligence in students' assessments: Applications, perceptions and implications in a Nigerian university. *Rima International Journal of Education*, 3(1), 205–217. ISSN: 2756-6749
- Moher, D., et al. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS medicine*, 6(7), e1000097.
- Oravec, J. (2022). AI, biometric analysis, and emerging cheating detection systems: The engineering of academic integrity? *Education Policy Analysis Archives*, 30(175).
- Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, 372, n71

- Popay, J., et al. (2006). Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC methods programme. Preprint.
- Popenici, S., & Kerr, S. (2017). Artificial Intelligence in Education: The Future of Learning. *Preprint*.
- Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: a fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint arXiv:2205.01833.
- Suazo Galdames, I. (2024). Integration of Artificial Intelligence in Higher Education: Relevance for Inclusion and Learning. *SciComm*, 4(1), 1–12.
- Tomlinson, C. A. (2001). How to differentiate instruction in mixed-ability classrooms. ASCD.
- Tomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., Bond, M., & Koryakina, A. (2023). EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis [Computer software]. EPPI Centre Software. UCL Social Research Institute, University College London.
- Wasif, S., & Hinton, G. (2024). Ethical and practical implications of AI in educational governance: A systematic analysis. Department of Computer Science, University of Colophonian.
- Williamson, B. (2019). Artificial Intelligence in Education: The Promise, the Potential, and the Pitfalls. *Preprint*.