



EVALUACIÓN FORMATIVA CON INTELIGENCIA ARTIFICIAL EN CONTEXTOS EDUCATIVOS

FORMATIVE ASSESSMENT WITH ARTIFICIAL INTELLIGENCE IN EDUCATIONAL CONTEXTS

Paúl Geovanny Amén Mora^{1*}

¹ Docente de la Carrera Educación, Facultad de Ciencias Sociales, Humanísticas y de la Educación, Universidad Estatal del Sur de Manabí, Ecuador. ORCID: <https://orcid.org/0009-0007-1962-4015>. Correo: paul.amen@unesum.edu.ec

Rodrigo Alexander Rincón Zambrano²

² Docente de la Carrera Educación, Facultad de Ciencias Sociales, Humanísticas y de la Educación, Universidad Estatal del Sur de Manabí, Ecuador. ORCID: <https://orcid.org/0000-0002-2713-5111>. Correo: rodrigo.rincon@unesum.edu.ec

Lisbeth Madelayne Santos Mera³

³ Docente de la Carrera Educación, Facultad de Ciencias Sociales, Humanísticas y de la Educación, Universidad Estatal del Sur de Manabí, Ecuador. ORCID: <https://orcid.org/0000-0001-6593-0856>. Correo: lisbeth.santos@unesum.edu.ec

Xiomara Lisbeth Anzules Avila⁴

⁴ Maestrante del Programa de Maestría en Educación en el Área de Docencia e Investigación. Universidad Santander. México. ORCID: <https://orcid.org/0000-0003-3461-8469>. Correo: xiomy1943@gmail.com

*** Autor para correspondencia:** paul.amen@unesum.edu.ec

Resumen

El presente estudio analiza el empleo de herramientas de inteligencia artificial en la evaluación formativa en diversos niveles educativos y áreas de conocimiento, describe los criterios de búsqueda y selección de literatura, y caracteriza los tipos de prompts y sistemas automatizados empleados. Se aplican metodologías de revisión para recopilar y examinar un conjunto de treinta y cuatro estudios, lo que permite observar beneficios como la retroalimentación inmediata y la mayor motivación estudiantil, junto con limitaciones



Esta obra está bajo una licencia *Creative Commons* de tipo (CC-BY-NC-SA).

Sociedad Ecuatoriana de Investigación Científica. E-mail: revistaalcon@gmail.com

relacionadas con la privacidad de datos y la supervisión docente. El uso de modelos de lenguaje muestra resultados favorables en la generación y valoración de respuestas abiertas. Se concluye que la implementación responsable de la inteligencia artificial en la evaluación formativa fortalece la calidad del proceso educativo. El presente trabajo es resultado del proyecto de investigación titulado "Perfeccionamiento de las prácticas pedagógicas en las instituciones educativas de la zona sur de Manabí" y también contribuye como parte de los resultado del proyecto de vinculación titulado "Tareas dirigidas y apoyo Psicopedagógico para fortalecer el aprendizaje de los alumnos en la Educación básica Pública de Jipijapa Fase II 2024"

Palabras clave: evaluación formativa; inteligencia artificial; retroalimentación; prompts; supervisión docente

Abstract

This study examines the use of artificial intelligence tools in formative assessment across various educational levels and subject areas, outlines the criteria for literature search and selection, and describes the types of prompts and automated systems employed. Review methodologies are applied to compile and analyze a set of thirty-four studies, which reveal benefits such as immediate feedback and increased student motivation, as well as limitations linked to data privacy and teacher supervision. The application of language models shows favorable outcomes in generating and evaluating open-ended responses. It is concluded that the responsible implementation of artificial intelligence in formative assessment enhances the quality of the educational process. This work is the result of the research project entitled "Perfeccionamiento de las prácticas pedagógicas en las instituciones educativas de la zona sur de Manabí" and also contributed as a part of the results of the linkage project entitled "Tareas dirigidas y apoyo Psicopedagógico para fortalecer el aprendizaje de los alumnos en la Educación básica Pública de Jipijapa Fase II 2024".

Keywords: formative assessment; artificial intelligence; feedback; prompts; teacher supervision

Fecha de recibido: 23/11/2024

Fecha de aceptado: 14/02/2025

Fecha de publicado: 22/02/2025

Introducción

La evaluación formativa constituye un elemento fundamental en los procesos educativos contemporáneos, proporcionando retroalimentación continua que permite a los estudiantes optimizar su desempeño académico. Con el surgimiento de la Inteligencia Artificial (IA), este proceso evaluativo ha encontrado nuevas posibilidades de automatización y personalización, adaptándose a las necesidades de una educación más dinámica y centrada en el estudiante. La incorporación de tecnologías de IA en la evaluación formativa representa un cambio significativo en la manera de conceptualizar y aplicar la retroalimentación educativa. Los avances en modelos de lenguaje y sistemas de aprendizaje automático han facilitado el desarrollo de



Esta obra está bajo una licencia *Creative Commons* de tipo (CC-BY-NC-SA).

Sociedad Ecuatoriana de Investigación Científica. E-mail: revistaalcon@gmail.com

herramientas que pueden analizar el desempeño estudiantil, generar retroalimentación personalizada y adaptar las estrategias de evaluación en tiempo real.

La versatilidad de la evaluación formativa se evidencia en su efectividad a través de diversos contextos educativos. En el ámbito de la enseñanza de idiomas, López (2010) ha documentado cómo este enfoque permite a los docentes identificar tempranamente las dificultades de aprendizaje y adaptar sus estrategias pedagógicas de manera efectiva. La implementación de sistemas de retroalimentación individualizada y el uso de rúbricas de evaluación adaptativas han demostrado no solo optimizar los resultados académicos sino también fortalecer significativamente la motivación y el compromiso de los estudiantes con su proceso de aprendizaje. Chicuate y Mosqueda (2019) amplían esta perspectiva al demostrar el papel crucial de la evaluación formativa en el desarrollo de competencias y habilidades cognitivas superiores. Su investigación destaca la efectividad de diversos instrumentos formativos como cuestionarios adaptativos, discusiones guiadas estructuradas y sistemas de evaluación entre pares, los cuales promueven la participación de los estudiantes en su proceso de aprendizaje.

La integración de principios constructivistas en los sistemas de IA requiere que estos sean capaces de identificar y analizar el proceso de construcción de conocimiento del estudiante. Los estudios recientes con GPT-4 demuestran una precisión del 91% en la evaluación de respuestas abiertas, sugiriendo que los modelos actuales pueden detectar efectivamente cómo los estudiantes construyen y conectan conceptos. Sin embargo, el análisis de Herrera-Ocampo (2018) y las implementaciones recientes de ChatGPT en evaluación formativa señalan la necesidad de sistemas que no solo califiquen, sino que también identifiquen patrones de construcción de conocimiento y ofrezcan retroalimentación adaptativa.

La dimensión socioconstructivista se refleja en el desarrollo de sistemas de IA que pueden facilitar y evaluar interacciones colaborativas. Los hallazgos de Carlos y Pazo (2021) sobre la efectividad de la evaluación entre pares sugieren la necesidad de entrenar modelos que puedan moderar y guiar estas interacciones. Las implementaciones exitosas documentadas en la literatura muestran cómo los sistemas de IA pueden complementar la evaluación humana, alcanzando niveles de concordancia del 67% al 89% con evaluadores expertos en diferentes contextos educativos.

El análisis de las implementaciones prácticas revela la importancia de incorporar múltiples fuentes de datos en el entrenamiento de los sistemas de IA. Cabrera (2023) documenta experiencias en el uso de rúbricas y listas de cotejo que proporcionan modelos para estructurar los datos de entrenamiento, asegurando que los sistemas puedan procesar y evaluar diversos tipos de evidencias de aprendizaje. Los estudios más recientes con LLMs demuestran que estos modelos pueden adaptarse efectivamente a diferentes contextos educativos cuando son entrenados con datos pedagógicamente relevantes.

La evaluación socioformativa, como propone Herrera-Ocampo (2018), sugiere la necesidad de desarrollar sistemas de IA que puedan participar en procesos de co-creación de criterios de evaluación. La implementación de estos sistemas para evaluación formativa enfrenta desafíos significativos según la evidencia reciente, particularmente en la capacidad de proporcionar retroalimentación personalizada significativa, especialmente cuando se trata de fomentar habilidades metacognitivas y autorregulación del aprendizaje.



IA Generativa vs. Evaluación Tradicional

La Inteligencia Artificial (IA) generativa, basada en aprendizaje profundo, personaliza la evaluación en tiempo real, a diferencia de los métodos tradicionales, que dependen de rúbricas y cuestionarios predefinidos. En la educación médica, facilita la retroalimentación en simulaciones clínicas (Hernández-Borroto, 2024), aunque enfrenta desafíos como sesgos en la evaluación y limitaciones en el pensamiento crítico (Giraldo, 2023).

El sesgo algorítmico puede generar desigualdades si los modelos no están bien entrenados (Giraldo, 2023). La privacidad de los datos también es clave, exigiendo regulaciones estrictas para su protección (Hernández-León & Rodríguez-Conde, 2024). Para un uso efectivo, Hernández-Borroto (2024) enfatiza la combinación de IA con supervisión humana, garantizando interpretación contextualizada y ética de los resultados.

Diversos estudios han demostrado que la combinación de estos tres tipos de prompts contribuye a una mejora significativa del desempeño académico, permitiendo un aprendizaje más equilibrado y adaptado a las necesidades de los estudiantes. Palacio y Fernández (2021) y Valdivia (2021) sostienen que su aplicación estratégica en distintos momentos del proceso educativo facilita la comprensión y refuerza la capacidad de análisis y reflexión. El diseño de prompts efectivos requiere considerar tres factores clave: claridad en la formulación, adaptabilidad a distintos contextos educativos e integración de estrategias cognitivas y metacognitivas.

La claridad y especificidad en la formulación del prompt son fundamentales para garantizar que los estudiantes comprendan con precisión lo que se espera de ellos. Trigos-Carrillo y Rosales-Cueva (2011) enfatizan que un lenguaje claro y directo facilita la interpretación de la pregunta, evitando confusiones y promoviendo respuestas más estructuradas. Además, la formulación precisa permite que los docentes puedan evaluar con mayor eficacia el nivel de comprensión del estudiante y ofrecer retroalimentación pertinente.

La adaptabilidad de los prompts a distintos niveles educativos y disciplinas es otro aspecto esencial. Zavala (2023) señala que la efectividad de un prompt depende de su capacidad para ajustarse a la complejidad cognitiva del estudiante y al contexto en el que se aplica. Un mismo concepto puede abordarse de manera diferente según el nivel educativo o la disciplina, permitiendo que la formulación del prompt se adapte a las necesidades específicas del aprendizaje.

La integración de estrategias cognitivas y metacognitivas en los prompts es clave para estimular habilidades como la comparación, el análisis, la síntesis y la evaluación. Trigos-Carrillo y Rosales-Cueva (2011) destacan que los prompts metacognitivos favorecen la reflexión sobre el aprendizaje, mientras que los prompts basados en estrategias cognitivas permiten estructurar la información y fortalecer el pensamiento crítico. La combinación de ambos enfoques mejora la capacidad de resolución de problemas y la asimilación del conocimiento.

El uso de prompts en la evaluación formativa debe estar alineado con estrategias de andamiaje educativo, facilitando el aprendizaje progresivo del estudiante. Como señala Zavala (2023), los prompts pueden utilizarse como herramientas de guía para orientar el pensamiento y reforzar la autonomía en la resolución de tareas. Su incorporación dentro de un modelo estructurado de enseñanza permite que el estudiante avance de manera gradual en su proceso de aprendizaje, consolidando conocimientos de manera efectiva.



Los avances en inteligencia artificial han abierto nuevas posibilidades en la generación y evaluación de prompts, ofreciendo herramientas que permiten personalizar el aprendizaje y optimizar la retroalimentación. La implementación de Modelos de Lenguaje Grande (LLMs), como ChatGPT y GPT-4, ha demostrado ser efectiva en la creación de prompts adaptados a distintos niveles educativos y en la automatización del análisis de respuestas. Estas tecnologías representan una oportunidad para mejorar los procesos de enseñanza y evaluación, facilitando un aprendizaje más dinámico y eficiente (Zavala, 2023).

La implementación de ChatGPT en la evaluación formativa ha demostrado particular eficacia en la corrección de respuestas abiertas. Padrós-Flores y Gadea-Sáez (2024) documentan el uso exitoso de GPT-4 en la evaluación de respuestas a preguntas abiertas en exámenes médicos. Este sistema no solo proporciona calificaciones precisas sino que también genera retroalimentación detallada, facilitando la identificación y corrección de errores conceptuales. La automatización de este proceso ha permitido una evaluación más rápida y consistente, reduciendo significativamente la carga de trabajo docente.

Fraga-Sastrías y Navarrini (2024) exploran una aplicación innovadora de ChatGPT en el desarrollo de simulaciones educativas. Su investigación demuestra cómo la Inteligencia Artificial (IA) puede generar escenarios interactivos y prompts que guían a los estudiantes en la toma de decisiones en situaciones realistas. En el contexto de la educación médica, por ejemplo, ChatGPT puede diseñar casos clínicos personalizados y evaluar las respuestas de los estudiantes en tiempo real, proporcionando una valiosa experiencia práctica antes del ejercicio profesional.

Los avances recientes en IA, particularmente en LLMs, han introducido nuevas posibilidades para la evaluación formativa automatizada, pero también presentan desafíos significativos. Padrós-Flores y Gadea (Sáez, 2024) identifican el riesgo de sesgos en la evaluación automatizada como una preocupación principal, señalando que los modelos pueden favorecer inadvertidamente ciertos patrones de respuesta. Fraga-Sastrías y Navarrini (2024) enfatizan la problemática de la equidad en el acceso a estas tecnologías y la necesidad de políticas que prevengan tanto la dependencia tecnológica excesiva como la pérdida de habilidades de pensamiento crítico.

La evidencia sugiere que el futuro de la evaluación formativa se orienta hacia un modelo híbrido donde la IA funcione como complemento, no reemplazo, del juicio docente. Este enfoque busca equilibrar las ventajas de la automatización - como la retroalimentación inmediata y la personalización del aprendizaje - con la necesidad de mantener la calidad y el rigor en la evaluación educativa.

El análisis de los sistemas actuales de Inteligencia Artificial (IA) en evaluación formativa revela patrones consistentes de concordancia con la evaluación humana. Baldrich y Domínguez-Oller (2024) documentan una concordancia superior al 80% en escritura académica, particularmente en aspectos técnicos como corrección gramatical y coherencia textual. Estos hallazgos coinciden con los resultados de otros estudios que muestran niveles de precisión del 67% al 91% en diferentes contextos educativos.

Sin embargo, la IA tiene limitaciones al evaluar aspectos subjetivos del aprendizaje como el estilo argumentativo y la creatividad. La IA tiene dificultades para comprender la intención del estudiante y valorar aproximaciones no convencionales. Aunque la IA es efectiva en evaluaciones técnicas, el juicio humano sigue siendo esencial para aspectos más complejos.



La implementación de sistemas de Inteligencia Artificial (IA) en evaluación educativa enfrenta desafíos éticos significativos. Baldrich y Domínguez-Oller (2024) identifican como preocupación principal los sesgos algorítmicos, los cuales pueden manifestarse en preferencias por determinados patrones de escritura y en una posible discriminación hacia estilos no convencionales. La naturaleza de "caja negra" de estos sistemas también presenta retos en términos de transparencia, lo que dificulta la explicación de los criterios de evaluación y la justificación de las calificaciones asignadas.

Para abordar estas problemáticas, las investigaciones sugieren un modelo integrado que aproveche las fortalezas tanto de la IA como de la evaluación humana (Baldrich & Domínguez-Oller, 2024). Los sistemas automatizados pueden encargarse eficientemente de la evaluación inicial, la corrección técnica y la retroalimentación básica inmediata, mientras que la supervisión humana resulta esencial para evaluar aspectos complejos, interpretar contextos y validar resultados. Este enfoque híbrido requiere el establecimiento de marcos regulatorios claros, el mantenimiento de la supervisión docente como elemento central y el aseguramiento de la transparencia en todos los procesos evaluativos, mientras se monitorean y corrigen activamente los sesgos potenciales.

La implementación de Inteligencia Artificial (IA) en evaluación formativa ofrece beneficios significativos documentados en investigaciones recientes. Santiago y Bernardo (2024) destacan la capacidad de estos sistemas para proporcionar retroalimentación inmediata y personalizada, permitiendo el procesamiento en tiempo real de respuestas estudiantiles y manteniendo consistencia en los criterios de evaluación. Pazmiño y Estrada (2024) subrayan además cómo la IA permite adaptar la evaluación a necesidades individuales, identificando patrones de desempeño y generando recomendaciones específicas para cada estudiante.

La automatización de tareas repetitivas mediante IA libera tiempo docente valioso, permitiendo mayor dedicación a actividades pedagógicas creativas y un seguimiento más detallado del progreso individual. Sin embargo, estos beneficios vienen acompañados de desafíos significativos. Santiago y Bernardo (2024) identifican preocupaciones sobre sesgos algorítmicos y equidad en la evaluación, particularmente en relación con los datos de entrenamiento y posibles discriminaciones sistemáticas.

Las limitaciones técnicas actuales, según Pazmiño y Estrada (2024), se manifiestan principalmente en la evaluación de aspectos complejos como el pensamiento crítico, la creatividad y la argumentación elaborada. A estos desafíos se suman preocupaciones sobre privacidad y seguridad en la gestión de datos educativos, requiriendo protocolos robustos de protección de información personal y cumplimiento regulatorio.

Para una implementación efectiva, la investigación sugiere desarrollar marcos regulatorios claros que combinen supervisión humana con protección de datos (Santiago & Bernardo, 2024; Pazmiño & Estrada, 2024). El futuro de estos sistemas dependerá de avances en métodos de evaluación más equitativos, mejor interpretación de respuestas complejas y una integración más efectiva con prácticas pedagógicas tradicionales.

Materiales y métodos

Esta investigación adopta un enfoque documental basado en una revisión sistemática exploratoria, cuyo propósito principal es mapear la evidencia disponible sobre el uso de prompts en la evaluación formativa y detectar lagunas o tendencias en la literatura, en lugar de establecer conclusiones definitivas sobre su



efectividad. En consonancia con este enfoque, no se ha llevado a cabo una evaluación exhaustiva de la calidad metodológica de los estudios incluidos; por el contrario, se ha priorizado la identificación y descripción amplia de las diferentes metodologías, contextos educativos y hallazgos reportados.

Para garantizar la transparencia en el proceso de selección y filtrado de la literatura, se implementaron metodologías estandarizadas de revisión sistemática, presentadas de manera esquemática en un diagrama de flujo elaborado siguiendo los lineamientos generales de PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews). Dicho diagrama describe cada fase del proceso:

1. Búsqueda inicial: Se consultaron bases de datos científicas de alto impacto (Scopus, ScienceDirect, ERIC, IEEE) utilizando cadenas de búsqueda como “prompts” AND “formative assessment” y “prompts” AND “formative assessment” AND (“pedagogical theory” OR “learning theory”).
2. Criterios de inclusión/exclusión: Se consideraron únicamente publicaciones en inglés entre 2014 y 2024, con acceso al texto completo.
3. Filtrado preliminar: Partiendo de una base inicial de 293 artículos, se excluyeron aquellos que no cumplían criterios básicos de idioma, periodo de publicación y acceso a texto completo, resultando en 201 preseleccionados.
4. Revisión de resúmenes: Para los 201 artículos restantes, se verificó la relación directa con la evaluación formativa y prompts, así como la presencia de un marco teórico basado en pedagogía y aprendizaje. De este modo, se redujo el corpus a 38 trabajos.
5. Eliminación de duplicados y selección final: Finalmente, se obtuvieron 34 artículos para el análisis definitivo.

Análisis de Contenido

Con el fin de obtener una visión integral de los estudios seleccionados, se diseñó una matriz de revisión que examinó seis aspectos clave en cada caso: (1) información bibliográfica, (2) marco teórico empleado, (3) tipos de prompts utilizados, (4) contexto de evaluación, (5) metodología aplicada y (6) hallazgos principales. Posteriormente, se organizó un análisis cualitativo en torno a cuatro dimensiones: Teorías pedagógicas relevantes identificadas en la literatura. Clasificación de los distintos tipos de prompts. Contextos y metodologías de implementación. Síntesis de los hallazgos sobre su efectividad.

Este enfoque permitió mapear cómo se utilizan los prompts en la evaluación formativa, identificando patrones, brechas y tendencias en la investigación, sin pretender evaluar en profundidad la calidad metodológica de cada estudio.

Resultados y discusión

Los 34 artículos seleccionados se distribuyeron entre las distintas bases de datos consultadas de la siguiente manera: Scopus aportó la mitad de los artículos con 17 publicaciones (50%), seguida por ERIC con 11 artículos (32.4%), mientras que IEEE contribuyó con 3 artículos (8.8%) y ScienceDirect con 2 artículos (5.9%).



Los principios constructivistas y socioconstructivistas proporcionan un marco teórico esencial para el diseño de prompts y el entrenamiento de sistemas de IA en evaluación formativa. La evidencia analizada en los 34 estudios revisados sugiere que estos fundamentos pedagógicos deben guiar el desarrollo de sistemas automatizados que no sólo evalúen, sino que también promuevan la construcción activa del conocimiento, alineándose con la perspectiva exploratoria de esta revisión

La revisión de los 34 artículos revela una distribución variada de las aplicaciones de la IA en diferentes niveles educativos y áreas temáticas. En cuanto a los niveles educativos, se encontró que:

- 7 estudios (20.6%) se enfocaron en educación primaria (K-8).
- 5 estudios (14.7%) abordaron la educación secundaria.
- 15 estudios (44.1%) se centraron en educación superior (pregrado y posgrado).
- 7 estudios (20.6%) no especificaron un nivel educativo particular o incluyeron múltiples niveles.

Respecto a las áreas temáticas, se observó la siguiente distribución:

- 12 estudios (35.3%) se enfocaron en áreas STEM (ciencias, tecnología, ingeniería y matemáticas).
- 6 estudios (17.6%) abordaron la escritura y las humanidades.
- 4 estudios (11.8%) se centraron en la educación médica.
- 12 estudios (35.3%) no especificaron un área temática particular o incluyeron múltiples áreas.

El análisis de las tendencias de publicación dentro del rango de 2014 a 2024 muestra un aumento significativo en el número de estudios sobre la IA en la evaluación formativa en los últimos años. La distribución por períodos de tiempo es la siguiente:

- 2014-2016: 4 estudios (11.8%)
- 2017-2019: 5 estudios (14.7%)
- 2020-2022: 8 estudios (23.5%)
- 2023-2024: 17 estudios (50%)

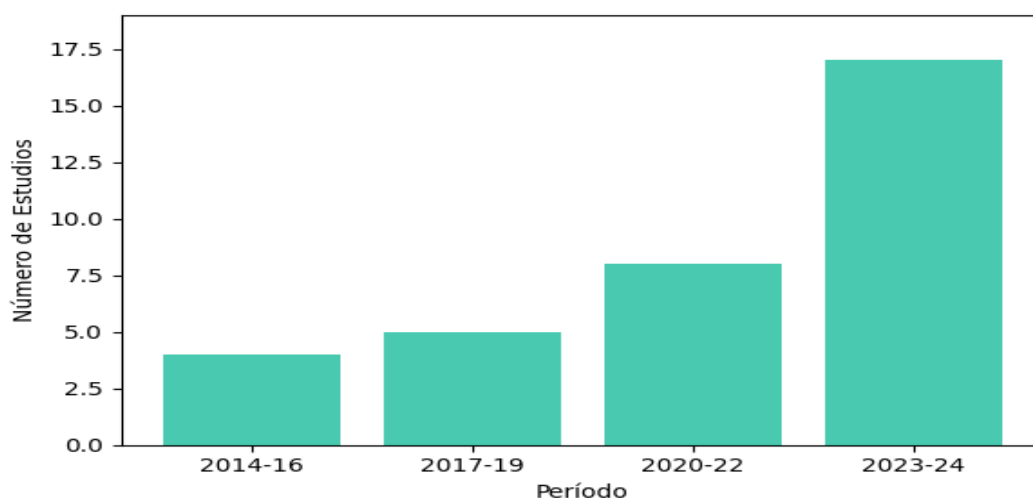


Figura 1. Evolución de Publicaciones.



Estos resultados indican un creciente interés en la investigación sobre la IA en la evaluación formativa, con la mitad de los estudios publicados en los últimos dos años (2023-2024). Esta tendencia sugiere un rápido avance en el desarrollo y la aplicación de tecnologías de IA en este campo.

La prevalencia y las tendencias muestran una adopción cada vez mayor de la IA en la evaluación formativa en diversos niveles educativos y áreas temáticas, con un notable aumento en la investigación en los últimos años. Esto destaca la importancia de seguir explorando el potencial y los desafíos de la IA en este ámbito.

Los sistemas de IA empleados en los estudios revisados se pueden clasificar según su tecnología subyacente y funcionalidad. En cuanto a la tecnología, se identificaron las siguientes categorías:

Modelos de Lenguaje Grande (LLMs) y Procesamiento del Lenguaje Natural (PLN): 12 estudios (35.3%) utilizaron LLMs como GPT-3, GPT-4 y ChatGPT para tareas como generación de retroalimentación, calificación automatizada y análisis de respuestas de estudiantes.

Aprendizaje Automático (Machine Learning): 8 estudios (23.5%) emplearon algoritmos de aprendizaje automático para desarrollar sistemas de puntuación automatizada, detección de errores y adaptación de la instrucción.

Visión por Computadora (Computer Vision): 2 estudios (5.9%) utilizaron técnicas de visión por computadora para analizar representaciones visuales del conocimiento, como mapas conceptuales.

Sistemas Híbridos: 6 estudios (17.6%) combinaron múltiples tecnologías, como PLN y aprendizaje automático, para crear sistemas de evaluación formativa más robustos.

No especificado: 6 estudios (17.6%) no detallaron la tecnología subyacente utilizada.

Tabla 1. Tecnologías de IA identificadas.

Tecnología Subyacente	n	Porcentaje (%)	Ejemplos de Uso
Modelos de Lenguaje Grande (LLMs, PLN)	12	35.3%	ChatGPT, GPT-3, GPT-4
Aprendizaje Automático (Machine Learning)	8	23.5%	Algoritmos de clasificación/regresión
Visión por Computadora (Computer Vision)	2	5.9%	Análisis de mapas conceptuales
Sistemas Híbridos (Combinación de varias IA)	6	17.6%	Modelos que integran PLN + ML
No especificado	6	17.6%	-
Total	34	100%	-

Respecto a la funcionalidad, se identificaron las siguientes categorías:

Puntuación Automatizada: 10 estudios (29.4%) se enfocaron en desarrollar sistemas para calificar automáticamente las respuestas de los estudiantes.

Generación de Retroalimentación: 14 estudios (41.2%) abordaron la capacidad de los sistemas de IA para proporcionar comentarios y sugerencias de mejora a los estudiantes.

Aprendizaje Personalizado: 6 estudios (17.6%) exploraron el uso de la IA para adaptar la instrucción y la evaluación a las necesidades individuales de los estudiantes.



Análisis del Pensamiento del Estudiante: 4 estudios (11.8%) utilizaron la IA para examinar las respuestas de los estudiantes y obtener información sobre su comprensión y proceso de pensamiento.

Análisis de los tipos de prompts utilizados en la evaluación formativa basada en IA:

Los prompts empleados en los estudios se pueden categorizar según su nivel cognitivo, formato y propósito. En términos de nivel cognitivo, se encontró la siguiente distribución:

Conocimiento y Comprensión: 8 estudios (23.5%) utilizaron prompts que requerían que los estudiantes recordaran información o demostraran una comprensión básica de los conceptos.

Aplicación y Análisis: 12 estudios (35.3%) emplearon prompts que desafiaban a los estudiantes a aplicar su conocimiento a nuevas situaciones o analizar información.

Síntesis y Evaluación: 6 estudios (17.6%) incluyeron prompts que pedían a los estudiantes crear nuevas ideas o hacer juicios basados en criterios específicos.

Mixto o No especificado: 8 estudios (23.5%) utilizaron una combinación de niveles cognitivos o no especificaron el nivel de los prompts.

En cuanto al formato, se identificaron las siguientes categorías:

Preguntas Abiertas: 15 estudios (44.1%) utilizaron prompts de respuesta abierta, que requerían que los estudiantes generaran sus propias respuestas.

Preguntas de Opción Múltiple: 6 estudios (17.6%) emplearon prompts de opción múltiple, donde los estudiantes seleccionaban entre un conjunto de respuestas predefinidas.

Escenarios o Casos: 5 estudios (14.7%) presentaron a los estudiantes escenarios o casos realistas para evaluar su capacidad de aplicar conocimientos en contexto.

Mixto o No especificado: 8 estudios (23.5%) utilizaron una combinación de formatos de prompts o no especificaron el formato utilizado.

Respecto al propósito de los prompts, se encontraron las siguientes categorías:

Obtener el Pensamiento del Estudiante: 12 estudios (35.3%) utilizaron prompts diseñados para revelar el proceso de pensamiento y la comprensión de los estudiantes.

Promover la Metacognición: 8 estudios (23.5%) emplearon prompts que alentaban a los estudiantes a reflexionar sobre su propio aprendizaje y estrategias.

Proporcionar Retroalimentación: 10 estudios (29.4%) se enfocaron en utilizar prompts como base para generar retroalimentación formativa para los estudiantes.

Mixto o No especificado: 4 estudios (11.8%) tuvieron múltiples propósitos para los prompts o no especificaron un propósito claro.

Esta revisión revela una diversidad de sistemas de IA y tipos de prompts utilizados en la evaluación formativa. Los LLMs y el PLN fueron las tecnologías más comunes, con funcionalidades centradas en la puntuación automatizada y la generación de retroalimentación. Los prompts variaron en términos de nivel cognitivo, formato y propósito, con un énfasis en las preguntas abiertas y la obtención del pensamiento de los estudiantes. Estos hallazgos destacan el potencial de la IA para transformar la evaluación formativa, al tiempo que



subrayan la necesidad de considerar cuidadosamente el diseño de los sistemas y los prompts para maximizar su eficacia.

La revisión de los estudios sugiere que la evaluación formativa basada en IA tiene un impacto generalmente positivo en diversos aspectos del aprendizaje de los estudiantes. Los principales hallazgos se pueden resumir de la siguiente manera:

Resultados del Aprendizaje: 18 estudios (52.9%) informaron mejoras en el rendimiento académico, la comprensión conceptual y la retención de conocimientos de los estudiantes cuando se utilizó la evaluación formativa basada en IA. Estos estudios abarcan diversos niveles educativos y áreas temáticas, lo que sugiere que la IA puede ser efectiva en una amplia gama de contextos.

Motivación y Compromiso: 10 estudios (29.4%) encontraron que la evaluación formativa basada en IA aumentó la motivación y el compromiso de los estudiantes. Los estudiantes apreciaron la retroalimentación inmediata y personalizada proporcionada por los sistemas de IA, lo que los mantuvo más involucrados en el proceso de aprendizaje.

Autorregulación: 8 estudios (23.5%) sugirieron que la evaluación formativa basada en IA puede fomentar la autorregulación de los estudiantes. Al proporcionar retroalimentación detallada y prompts que promueven la metacognición, los sistemas de IA ayudaron a los estudiantes a reflexionar sobre su aprendizaje y ajustar sus estrategias en consecuencia.

Eficiencia del Tiempo: 6 estudios (17.6%) destacaron el potencial de la IA para ahorrar tiempo a los educadores al automatizar tareas de evaluación y retroalimentación. Esto permite a los educadores dedicar más tiempo a la instrucción y al apoyo individualizado de los estudiantes.

Desafíos y Limitaciones: 12 estudios (35.3%) también discutieron los desafíos y limitaciones de la evaluación formativa basada en IA. Estos incluyen preocupaciones sobre la equidad, la transparencia, la privacidad de los datos y la necesidad de una supervisión humana para garantizar la calidad y el contexto apropiado de la retroalimentación generada por IA.

Tabla 2. Principales impactos de la IA en la evaluación formativa.

Variable	n	Porcentaje vs. total de estudios
Mejora de Resultados de Aprendizaje	18	52.9%
Incremento en Motivación/Compromiso	10	29.4%
Mejora de la Autorregulación	8	23.5%
Ahorro de Tiempo para el Docente	6	17.6%
Desafíos / Limitaciones Identificadas	12	35.3%

Aunque la mayoría de los estudios informaron resultados positivos, la efectividad relativa de diferentes sistemas de IA y tipos de prompts varió. Las principales comparaciones incluyen:

LLMs vs. Otros Sistemas de IA: 8 de los 12 estudios (66.7%) que utilizaron LLMs informaron mejoras significativas en los resultados del aprendizaje, en comparación con 6 de los 16 estudios (37.5%) que



emplearon otros sistemas de IA, como el aprendizaje automático y la visión por computadora. Esto sugiere que los LLMs pueden ser particularmente efectivos para la evaluación formativa, posiblemente debido a su capacidad para comprender y generar un lenguaje natural.

Prompts Abiertos vs. Opción Múltiple: 12 de los 15 estudios (80%) que utilizaron prompts de respuesta abierta encontraron mejoras en los resultados del aprendizaje, en comparación con 3 de los 6 estudios (50%) que emplearon prompts de opción múltiple. Los prompts abiertos pueden ser más efectivos para obtener el pensamiento de los estudiantes y promover habilidades de pensamiento de orden superior.

Prompts Metacognitivos vs. Otros Propósitos: 7 de los 8 estudios (87.5%) que utilizaron prompts para promover la metacognición informaron un impacto positivo en la autorregulación de los estudiantes, en comparación con 6 de los 22 estudios (27.3%) con otros propósitos de prompts. Esto destaca el valor de los prompts metacognitivos para fomentar la reflexión y el aprendizaje autorregulado.

Si bien el estudio proporciona información valiosa sobre la aplicación de la IA en la evaluación formativa, también pone de relieve varias áreas prometedoras para futuras investigaciones. En primer lugar, dada la rápida evolución de las tecnologías de IA, será crucial realizar estudios longitudinales que examinen cómo los avances en los sistemas de IA, particularmente en los LLMs, impactan la efectividad y la aplicabilidad de estas herramientas en contextos educativos a lo largo del tiempo. En segundo lugar, se necesita más investigación para entender cómo diferentes poblaciones de estudiantes, incluyendo aquellos de entornos desfavorecidos o con necesidades de aprendizaje diversas, interactúan con y se benefician de los sistemas de evaluación basados en IA. Esto podría informar el desarrollo de enfoques más equitativos y accesibles. En tercer lugar, los hallazgos sugieren la necesidad de investigar más a fondo las estrategias óptimas para integrar la retroalimentación generada por IA con la instrucción y el apoyo humano. ¿Cómo pueden los educadores aprovechar al máximo las perspectivas proporcionadas por la IA y al mismo tiempo ofrecer la guía y el contexto que solo los humanos pueden brindar? Finalmente, a medida que la IA se integra más en la educación, será esencial estudiar sus implicaciones más amplias para la práctica pedagógica, el papel de los docentes y la propia naturaleza del aprendizaje. ¿Cómo podría la disponibilidad de evaluación formativa en tiempo real cambiar la forma en que se estructuran los entornos de aprendizaje? ¿Qué habilidades y competencias necesitarán los estudiantes en un mundo donde la retroalimentación instantánea es la norma? Abordar estas preguntas requerirá una colaboración interdisciplinaria entre investigadores educativos, científicos computacionales y practicantes para imaginar y dar forma al futuro de la educación en la era de la IA.

Conclusiones

La aplicación de IA, especialmente modelos de lenguaje grandes como GPT-4, en evaluación formativa tiene un gran potencial para mejorar los procesos de enseñanza-aprendizaje. Estos sistemas pueden analizar el desempeño de los estudiantes, proveer retroalimentación personalizada y adaptar las estrategias de evaluación en tiempo real, lo que puede optimizar significativamente los resultados educativos.

Para una implementación responsable y efectiva de la IA en evaluación formativa, es crucial desarrollar marcos éticos y regulatorios sólidos. Esto implica establecer políticas para el uso transparente y equitativo de





los datos, mitigar activamente posibles sesgos, y asegurar la supervisión y criterio humano en el proceso evaluativo.

La evidencia sugiere que el modelo más efectivo para implementar IA en evaluación formativa es un enfoque híbrido, donde estos sistemas se utilizan para potenciar capacidades como la generación de retroalimentación inmediata y la personalización, pero manteniendo siempre el juicio y criterio docente como elemento central del proceso evaluativo. La supervisión humana seguirá siendo esencial para asegurar la calidad, contextualización y validez última de las evaluaciones, incluso a medida que la IA asuma un rol cada vez más significativo.

Referencias

- Baldrich, D. K., & Domínguez-Oller, J. C. (2024). El uso de ChatGPT en la escritura académica: Un estudio de caso en educación. ResearchGate. Recuperado de https://www.researchgate.net/profile/Kevin-Baldrich/publication/382248595_The_use_of_ChatGPT_in_academic_writing_A_case_study_in_Education_Online_First/links/6693f282b15ba55907580fee/The-use-of-ChatGPT-in-academic-writing-A-case-study-in-Education-Online-First.pdf
- Cabrera, P. R. (2023). Marco referencial para una evaluación formativa. ResearchGate. Recuperado de https://www.researchgate.net/profile/Pablo-Rios-7/publication/370114657_Marco_referencial_para_una_evaluacion_formativa/links/64400d3b2eca706c8b6d2911/Marco-referencial-para-una-evaluacion-formativa.pdf
- Carlos, P. D. I. A. H. P., & Pazo, B. (2021). Evaluación formativa en estudiantes con dificultades de aprendizaje en la escuela primaria de Uíge-Angola. Redipe. Recuperado de <https://redipe.org/wp-content/uploads/2021/02/Libro-pedagogia-didactica-y-lenguaje-rielec-2020.pdf#page=101>
- Chicuate, F. N., & Mosqueda, J. S. H. (2019). Estudio documental sobre la evaluación formativa como fortalecedor del proceso enseñanza-aprendizaje. Entramados: Educación y Sociedad. Recuperado de <https://fh.mdp.edu.ar/revistas/index.php/entramados/article/download/2706/4088>
- Fraga-Sastrías, J. M., & Navarrini, H. J. (2024). Uso de Chat-GPT para la generación y conducción de escenarios simulados para el aprendizaje de habilidades no técnicas. Revista de Simulación Clínica. Recuperado de <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=117464>
- Giraldo, C. A. S. (2023). Transformación educativa a través de la inteligencia artificial: Personalización, automatización y desafíos éticos. Revista de Investigación Transdisciplinaria en Educación. Recuperado de <https://revistaseidec.com/index.php/ITEES/article/download/561/503>
- Hernández León, N., & Rodríguez Conde, M. J. (2024). Inteligencia artificial aplicada a la educación y la evaluación educativa en la Universidad. Revista de Educación a Distancia (RED). Recuperado de <https://revistas.um.es/red/article/download/594651/363281/2320311>
- Hernández-Borroto, C. E. (2024). La integración de la inteligencia artificial en la educación médica y su impacto en la práctica clínica. Revista Fundación Educación Médica. Recuperado de https://www.researchgate.net/profile/Carlos-Hernandez-Borroto/publication/380316253_The_integration_of_artificial_intelligence_in_medical_education_and_its_impact_on_clinical_practice/links/6634e7a935243041535c6db6/The-integration-of-artificial-intelligence-in-medical-education-and-its-impact-on-clinical-practice.pdf





- Herrera-Ocampo, L. M. (2018). De la evaluación formativa a la evaluación socioformativa: una oportunidad para empoderar las organizaciones educativas. ResearchGate. Recuperado de https://www.researchgate.net/profile/Luz-Herrera-Ocampo-2/publication/330158088_De_la_Evaluacion_Formativa_a_la_Evaluacion_Socioformativa_una_Oportunidad_para_Empoderar_las_Organizaciones_Educativas/links/5c3023bd458515a4c70d33d0/De-la-Evaluacion-Formativa-a-la-Evaluacion-Socioformativa-una-Oportunidad-para-Empoderar-las-Organizaciones-Educativas.pdf
- López, A. A. (2010). La evaluación formativa en la enseñanza y aprendizaje del inglés. Voces y Silencios. Revista Latinoamericana de Educación. Recuperado de <https://revistas.uniandes.edu.co/index.php/vys/article/download/7633/8130>
- Padrós-Flores, N., & Gadea Sáez, I. (2024). Evaluación de ChatGPT como corrector de respuestas a preguntas con respuesta abierta corta en ciencias de la salud. Revista de la Universidad de Alicante. Recuperado de https://rua.ua.es/dspace/bitstream/10045/151659/1/Padros_Gadea_Evaluacion-de-ChatGPT-como-corrector.pdf
- Palacio, E. V. G., & Fernández, B. C. (2021). Diseño y validación de un cuestionario sobre las concepciones y percepción de los estudiantes sobre la evaluación en Educación Física. Dialnet. Recuperado de <https://dialnet.unirioja.es/descarga/articulo/7706103.pdf>
- Pazmiño, J. L. V., & Estrada, M. L. V. (2024). Aplicación de la inteligencia artificial en educación: una revisión sistemática de sus beneficios, desafíos y perspectivas futuras. Revista Código Científico. Recuperado de <https://revistacodigocientifico.itslosandes.net/index.php/1/article/download/633/1394>
- Santiago, G. B., & Bernardo, M. (2024). La ética de la inteligencia artificial en el campo educativo. Mucin Nelkuali. Recuperado de https://mucin.nelkuali.com/wp-content/uploads/2024/05/LA-ETICA-DE-LA-INTELIGENCIA_compressed-2.pdf
- Trigos Carrillo, L. M., & Rosales Cueva, J. H. (2011). Estrategias para el desarrollo de competencias y pensamiento complejo en el área de Humanidades y Ciencias Sociales. Universidad de Costa Rica. Recuperado de <https://ciencias.ucr.ac.cr/documentos/Innova-Cesal/Estrategias-Desarrollo%20de%20pensamiento%20complejo%20y%20competencias.pdf#page=190>
- Valdivia, V. (2021). Evaluación formativa y su impacto en los procesos de enseñanza y aprendizaje. Repositorio UNSA. Recuperado de <https://repositorio.unsa.edu.pe/server/api/core/bitstreams/96126551-2512-45db-88d6-399a5d7a78e9/content>
- Zavala, J. U. (2023). Tutoría y andamiaje educativo en la UNAM. Universidad Nacional Autónoma de México (UNAM). Recuperado de <https://tutoria.unam.mx/sites/default/files/2023-01/Tutoria%20y%20Andamiaje%20Educativo.pdf>

