

DOI: <https://doi.org/10.70577/cieninter.v4i3.138>

Inteligencia Artificial generativa como herramienta de retroalimentación del aprendizaje: revisión integrativa de evidencia, oportunidades y desafíos

Generative Artificial Intelligence as a Learning Feedback Tool: An Integrative Review of Evidence, Opportunities, and Challenges

Cristian Alexander Herrera Cevallos¹

Universidad Estatal de Bolívar, Guaranda, Ecuador

craherre2020@gmail.com

<https://orcid.org/0009-0003-5458-8129>

Karen Paola Quiñónez Valencia²

Universidad Técnica Luis Vargas Torres, Esmeraldas, Ecuador

karen.quinonez.valencia@utelvt.edu.ec

<https://orcid.org/0009-0002-5305-2595>

Jenny Antonieta Méndez Vivar³

Universidad Técnica Luis Vargas Torres, Esmeraldas, Ecuador

Jenny.mendez@utelvt.edu.ec

<https://orcid.org/0000-0002-3167-6690>

Luis Alexander Cabezas Gallegos⁴

Universidad Internacional del Ecuador, Quito, Ecuador

lucabezasga@uide.edu.ec

<https://orcid.org/0009-0009-8526-7305>

Irina Jacqueline Carvajal Alcívar⁵

Universidad Central del Ecuador, Médico, Quito, Ecuador

jacqui.carvajal1223@gmail.com

<https://orcid.org/0009-0003-5303-4693>

Cómo citar

Herrera Cevallos, C. A., Quiñónez Valencia, K. P., Méndez Vivar, J. A., Cabezas Gallegos, L. A., & Carvajal Alcívar, I. J. (2026). Inteligencia Artificial generativa como herramienta de retroalimentación del aprendizaje: revisión integrativa de evidencia, oportunidades y desafíos. *Ciencia Interdisciplinaria Internacional*, 4(3), 1297–1311. <https://doi.org/10.70577/cieninter.v4i3.138>

Fecha de recepción: 2026-09-01

Fecha de aceptación: 2026-09-15

Fecha de publicación: 2026-09-29

RESUMEN

La incorporación de la Inteligencia Artificial generativa (IAGen) en educación ha generado una transformación progresiva en las prácticas de enseñanza, aprendizaje y evaluación, particularmente en aquellas actividades que requieren interacción continua, revisión de productos y retroalimentación formativa. En este contexto, los modelos de lenguaje de gran escala han adquirido relevancia debido a su capacidad para analizar producciones académicas y proporcionar respuestas personalizadas en tiempo reducido. El objetivo de este artículo fue analizar críticamente la evidencia disponible sobre el uso de Inteligencia Artificial generativa como herramienta de retroalimentación del aprendizaje, con énfasis en sus efectos sobre el desempeño académico, la revisión de producciones, la autorregulación y la calidad de la retroalimentación. Se desarrolló una revisión integrativa de literatura científica reciente, complementada con revisiones sistemáticas, metaanálisis y documentos internacionales sobre el uso educativo de la IA generativa. La síntesis evidencia resultados favorables, aunque heterogéneos. La investigación reciente muestra que la IAGen puede contribuir a mejorar la calidad de las revisiones, la precisión lingüística, la productividad escrita, la participación y determinados resultados de aprendizaje. Un metaanálisis de 68 estudios reportó un efecto positivo global de magnitud moderada sobre los resultados de aprendizaje ($SMD = 0,45$; $IC95\% [0,43; 0,47]$), acompañado de elevada heterogeneidad ($I^2 = 95\%$). Paralelamente, revisiones específicas sobre retroalimentación mediante modelos de lenguaje señalan que la IA presenta mayores fortalezas en identificación de errores y aspectos estructurales, mientras que la retroalimentación docente mantiene ventajas en cuestiones de argumentación, contextualización, priorización de mejoras y orientación dialógica. Se concluye que la IAGen posee mayor valor educativo cuando funciona como mecanismo complementario dentro de un modelo de retroalimentación híbrida, en el que la tecnología aporta inmediatez, disponibilidad y capacidad de iteración, mientras que el docente conserva la responsabilidad de contextualizar, validar y profundizar el proceso de aprendizaje.

Palabras clave: inteligencia artificial generativa; retroalimentación formativa; aprendizaje; evaluación educativa; modelos de lenguaje; educación superior; ChatGPT.

ABSTRACT

The incorporation of Generative Artificial Intelligence (GenAI) into education has progressively transformed teaching, learning, and assessment practices, particularly in activities requiring continuous interaction, product revision, and formative feedback. In this context, large language models have gained relevance because of their ability to analyze academic productions and provide personalized responses within short periods of time. This article aimed to critically analyze the available evidence on the use of Generative Artificial Intelligence as a learning feedback tool, with emphasis on its effects on academic performance, product revision, self-regulation, and feedback quality. An integrative review of recent scientific literature was conducted, complemented by systematic reviews, meta-analyses, and international reports addressing the educational use of generative AI. The synthesis shows generally favorable but heterogeneous results. Recent research indicates that GenAI can contribute to improving revision quality, linguistic accuracy, writing productivity, engagement, and selected learning outcomes. A meta-analysis of 68 studies reported a moderate overall positive effect of GenAI on learning outcomes ($SMD = 0.45$; $95\% CI [0.43, 0.47]$), accompanied by substantial heterogeneity ($I^2 = 95\%$). At the same time, reviews focused specifically on large language model-based feedback indicate that AI performs particularly well in identifying errors and structural aspects, whereas teacher feedback retains advantages for argument development, contextual interpretation, prioritization of revision needs, and dialogic guidance. It is concluded that GenAI has greater educational value when used as a complementary mechanism within a hybrid feedback model, in which technology contributes immediacy, availability, and iterative interaction, while teachers retain responsibility for contextualizing, validating, and deepening the learning process.

Keywords: generative artificial intelligence; formative feedback; learning; educational assessment; large language models; higher education; ChatGPT.

INTRODUCCIÓN

La transformación digital de la educación ha alcanzado una nueva etapa con la expansión de la Inteligencia Artificial generativa. A diferencia de tecnologías educativas anteriores, los actuales modelos de lenguaje de gran escala permiten establecer interacciones conversacionales, interpretar producciones escritas, formular explicaciones, generar preguntas y proporcionar sugerencias de revisión a partir de instrucciones expresadas en lenguaje natural. Esta capacidad ha trasladado el debate educativo desde la automatización de determinadas tareas hacia una cuestión de mayor complejidad: cómo integrar la inteligencia artificial en procesos pedagógicos sin debilitar la autonomía, el pensamiento crítico y la responsabilidad intelectual del estudiante.

La retroalimentación constituye uno de los espacios donde esta transformación adquiere especial relevancia. Su función educativa no consiste simplemente en comunicar al estudiante si una respuesta es correcta o incorrecta, sino en proporcionar información que permita comprender el desempeño alcanzado, identificar discrepancias respecto de los objetivos de aprendizaje y determinar acciones concretas para mejorar. Desde la perspectiva clásica de Hattie y Timperley (2007), la retroalimentación adquiere diferentes niveles de profundidad, desde la información centrada en la tarea hasta aquella relacionada con los procesos y la autorregulación. La literatura posterior ha mantenido esta distinción como una referencia importante para comprender por qué no toda retroalimentación produce los mismos efectos.

La irrupción de los modelos generativos plantea, por tanto, una cuestión sustantiva. Si una herramienta puede producir retroalimentación en segundos, ¿la disponibilidad de mayor cantidad de comentarios

implica necesariamente un mayor aprendizaje?

La evidencia disponible sugiere que la respuesta no puede reducirse a una relación directa entre tecnología y rendimiento. Una revisión sistemática de 129 estudios sobre retroalimentación asistida por inteligencia artificial encontró resultados generalmente favorables en percepciones, acciones y resultados de aprendizaje, pero también identificó limitaciones relacionadas con la transparencia de los sistemas y la necesidad de mejorar el reporte metodológico.

La literatura más reciente permite observar una segunda característica: los efectos dependen de las condiciones bajo las cuales se implementa la tecnología. Un metaanálisis de 68 estudios experimentales y cuasiexperimentales reportó un efecto positivo moderado de la IA generativa sobre los resultados de aprendizaje ($SMD = 0,45$), pero encontró una heterogeneidad muy elevada ($I^2 = 95\%$), lo cual indica que los efectos no son uniformes entre contextos, poblaciones, disciplinas y diseños de intervención.

Esta heterogeneidad resulta particularmente relevante en el ámbito de la retroalimentación. La IA puede identificar errores gramaticales, inconsistencias estructurales y determinados problemas de claridad con elevada rapidez. Sin embargo, la identificación de un error no equivale necesariamente a comprender la razón pedagógica del error ni a saber qué debe hacer el estudiante para desarrollar una competencia transferible.

Una revisión sistemática publicada en 2026 sobre modelos de lenguaje y retroalimentación formativa en escritura, que incluyó 34 estudios realizados en 19 países, encontró que los sistemas basados en LLM favorecen la productividad escrita, la precisión

lingüística, el compromiso y la alfabetización en retroalimentación. Sin embargo, el mismo estudio señala que la retroalimentación humana mantiene ventajas en aspectos de mayor complejidad, como el desarrollo argumentativo, la interpretación contextual, la priorización de necesidades de revisión y la orientación dialógica.

Por otra parte, una revisión sistemática de 2026 centrada en retroalimentación formativa generada por LLM sintetizó 47 estudios empíricos y concluyó que los sistemas pueden superar las condiciones sin retroalimentación y, bajo determinadas condiciones de diseño, aproximarse a los resultados de la retroalimentación docente. No obstante, también identificó riesgos recurrentes relacionados con alucinaciones, exceso de positividad y clasificación incorrecta de trabajos estudiantiles.

Estos resultados permiten identificar una tensión fundamental. La IA generativa puede resolver una parte importante del problema tradicional de la retroalimentación —su disponibilidad y oportunidad—, pero introduce nuevos desafíos relacionados con la precisión, la validación, la dependencia tecnológica y la calidad cognitiva de las interacciones.

En consecuencia, el problema científico no consiste en determinar si la IA generativa puede producir retroalimentación. Esta capacidad ha quedado demostrada. La cuestión de mayor importancia consiste en establecer qué tipo de retroalimentación puede producir, bajo qué condiciones resulta pedagógicamente útil y cómo debe articularse con la intervención docente para favorecer aprendizaje sostenible.

A partir de este problema, el objetivo del presente artículo es analizar críticamente la evidencia científica

reciente sobre el uso de la Inteligencia Artificial generativa como herramienta de retroalimentación del aprendizaje, identificando sus efectos, fortalezas, limitaciones y condiciones pedagógicas para una implementación responsable.

La retroalimentación formativa puede comprenderse como información proporcionada al estudiante con la finalidad de reducir la distancia entre su desempeño actual y el desempeño esperado. Su efectividad depende tanto de su contenido como de la capacidad del estudiante para interpretarla y utilizarla.

Hattie y Timperley (2007) plantearon un modelo estructurado alrededor de tres preguntas centrales: hacia dónde se dirige el aprendizaje, cómo se está desarrollando el desempeño y cuáles son los siguientes pasos que deberían adoptarse. Este modelo permite diferenciar la simple corrección de una retroalimentación orientada al aprendizaje.

La literatura educativa advierte que los comentarios centrados exclusivamente en la persona tienen menor utilidad que aquellos orientados a la tarea, el proceso y la autorregulación. Por ello, la incorporación de IA debe diseñarse para desplazar la interacción desde el simple señalamiento del error hacia la construcción de procesos de revisión y reflexión.

La Inteligencia Artificial generativa hace referencia a sistemas capaces de producir contenido nuevo a partir de patrones aprendidos en grandes conjuntos de datos. En educación, los modelos de lenguaje representan una de sus aplicaciones más visibles.

Su característica diferencial respecto de sistemas tradicionales de evaluación automatizada consiste en la posibilidad de mantener interacciones conversacionales y producir respuestas contextualizadas a partir de las entradas del usuario.

Sin embargo, la apariencia de contextualización no

garantiza comprensión pedagógica. Los modelos de lenguaje producen respuestas probabilísticas y pueden generar información incorrecta o recomendaciones poco apropiadas.

Por esta razón, UNESCO propone un enfoque centrado en las personas, que incorpore protección de datos, validación ética y pedagógica, equidad y protección de la agencia humana.

MATERIAL Y MÉTODOS

El estudio se desarrolló mediante una revisión integrativa de literatura científica, orientada a sintetizar evidencia proveniente de estudios empíricos, revisiones sistemáticas y metaanálisis sobre IA generativa y retroalimentación educativa.

La decisión metodológica de emplear una revisión integrativa responde a la heterogeneidad del campo. Las investigaciones disponibles utilizan diferentes diseños, niveles educativos, disciplinas, herramientas, medidas de resultado y condiciones de implementación. En este contexto, una síntesis exclusivamente cuantitativa de todos los estudios podría ocultar diferencias conceptuales importantes.

La revisión se estructuró siguiendo los principios de transparencia recomendados por PRISMA 2020 para la identificación, selección, evaluación y síntesis de literatura científica. PRISMA 2020 establece 27 elementos de reporte destinados a mejorar la transparencia de las revisiones sistemáticas.

Se priorizaron investigaciones que cumplieran los siguientes criterios:

- Abordar explícitamente Inteligencia Artificial generativa, ChatGPT, GPT o modelos de lenguaje.
- Relacionar la tecnología con

retroalimentación, evaluación formativa, revisión o mejora del aprendizaje.

- Incluir resultados educativos, procesos de revisión, rendimiento, autorregulación, escritura o calidad de la retroalimentación.
- Corresponder a estudios empíricos, revisiones sistemáticas o metaanálisis.
- Presentar información suficiente para interpretar los resultados.
- Priorizar publicaciones recientes, especialmente de 2024 a 2026.

Se excluyeron documentos exclusivamente técnicos, artículos de opinión sin evidencia empírica, publicaciones comerciales y estudios centrados únicamente en detección de plagio sin relación con retroalimentación.

La evidencia fue organizada en cuatro dimensiones:

- Efectos sobre resultados de aprendizaje.
- Calidad y características de la retroalimentación.
- Procesos de revisión y autorregulación.
- Riesgos y condiciones de implementación.

Debido a la heterogeneidad de las fuentes, los resultados se sintetizaron mediante análisis comparativo y temático. Los tamaños de efecto se conservaron únicamente cuando procedían directamente de estudios de síntesis previamente publicados.

RESULTADOS Y DISCUSIÓN

Evolución reciente de la investigación

La literatura especializada muestra un crecimiento acelerado del interés científico por los modelos de lenguaje y la retroalimentación educativa.

Tabla 1.

Evolución de estudios sobre LLM y retroalimentación formativa en escritura

Año	Número de estudios	Porcentaje
2023	2	5,9 %
2024	9	26,5 %
2025	15	44,1 %
2026	8	23,5 %
Total	34	100,0 %

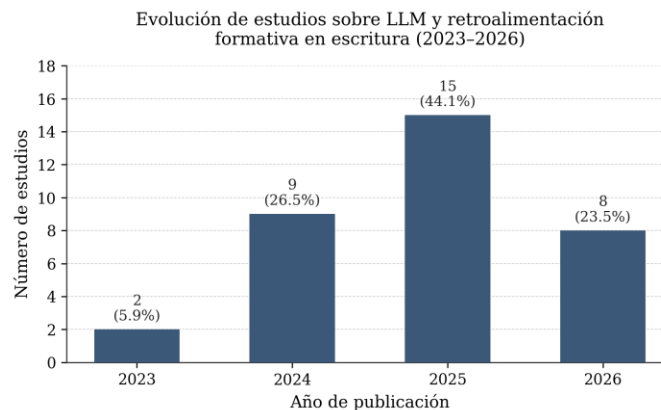
Nota. Elaboración propia a partir de los datos reportados por Jiao y Wang (2026). Los porcentajes fueron calculados sobre el total de 34 estudios incluidos en dicha revisión.

La distribución evidencia que aproximadamente siete de cada diez estudios incluidos en esta revisión fueron

publicados entre 2025 y 2026, lo que confirma el carácter emergente del campo.

Figura 1.

Evolución de la producción científica sobre LLM y retroalimentación formativa, 2023–2026.



La tendencia debe interpretarse como evidencia del crecimiento de la producción científica, no como demostración automática de efectividad. El aumento de publicaciones indica expansión del campo, pero no implica por sí mismo que los efectos educativos sean

cada vez mayores.

Magnitud de los efectos sobre el aprendizaje

Los estudios de síntesis disponibles muestran resultados favorables, aunque con una importante

variabilidad.

Han, Peng y Liu (2025) realizaron una revisión sistemática y metaanálisis de 68 estudios experimentales y cuasiexperimentales que produjeron 337 tamaños de efecto. El efecto global estimado fue

SMD = 0,45, con un intervalo de confianza del 95 % entre 0,43 y 0,47. Sin embargo, la heterogeneidad fue extremadamente elevada ($I^2 = 95\%$), indicando que el efecto depende de factores contextuales y metodológicos.

Tabla 2.

Evidencia cuantitativa reciente sobre IA generativa y aprendizaje

Estudio	Diseño	Evidencia sintetizada	Resultado principal	Interpretación
Han et al. (2025)	Revisión sistemática y metaanálisis	68 estudios; 337 tamaños de efecto	SMD = 0,45; $I^2 = 95\%$	Efecto global positivo moderado, con elevada heterogeneidad
Chen y Cheung (2025)	Revisión sistemática y metaanálisis	57 estudios; 97 estimaciones	$g = 0,804$	Efectos favorables en resultados académicos y pensamiento de orden superior; efecto no significativo en metacognición
Jiao y Wang (2026)	Revisión sistemática	34 estudios	Síntesis narrativa	Beneficios en escritura, precisión lingüística, productividad y alfabetización en retroalimentación
Maier et al. (2026)	Revisión sistemática cualitativa	47 estudios; 121 de preguntas investigación	Síntesis narrativa	Resultados favorables frente a ausencia de feedback; riesgos de alucinación, positividad excesiva y errores de clasificación
Lipnevich et al. (2026)	Scoping review	104 estudios empíricos	Síntesis comparativa	Alta variabilidad; modelos híbridos muestran resultados consistentes con la mediación humana

La comparación revela un elemento central: la existencia de un efecto promedio favorable no significa que la IA generativa produzca el mismo beneficio en cualquier situación educativa.

La elevada heterogeneidad encontrada por Han et al. (2025) constituye una advertencia metodológica relevante. Las diferencias entre nivel educativo, disciplina, duración de la intervención, tamaño muestral y diseño pedagógico pueden modificar

considerablemente los resultados.

Características de la retroalimentación generada por IA

La evidencia específica sobre feedback permite observar que las principales fortalezas de los modelos de lenguaje se concentran en aspectos relativamente estructurados.

Tabla 3.
Fortalezas y limitaciones de la retroalimentación generada por IA

Dimensión	Evidencia observada	Nivel de desarrollo
Corrección lingüística	Identificación de errores gramaticales y problemas de forma	Alto
Claridad	Sugerencias de reformulación y precisión	Alto
Estructura	Identificación de problemas organizativos	Alto
Revisión iterativa	Posibilidad de múltiples ciclos de interacción	Alto
Explicación	Capacidad para explicar determinados errores	Medio-alto
Argumentación	Evaluación de coherencia argumentativa compleja	Medio
Contextualización	Comprensión de circunstancias específicas del estudiante	Medio
Priorización	Determinación de qué problema debe corregirse primero	Medio-bajo
Autorregulación	Promoción autónoma de estrategias de aprendizaje	Variable
Juicio pedagógico	Interpretación integral del desempeño	Superior en modelos con intervención docente

Nota. Síntesis cualitativa basada principalmente en Jiao y Wang (2026), Maier et al. (2026) y Lipnevich et al. (2026).

La tabla muestra una tendencia consistente: la IA presenta mayores ventajas cuando la tarea de retroalimentación puede expresarse mediante criterios relativamente explícitos. A medida que aumenta la necesidad de interpretación contextual, juicio profesional y diálogo pedagógico, aumenta la importancia de la mediación humana.

Evidencia específica sobre escritura académica

Uno de los ámbitos con mayor desarrollo empírico

corresponde a la escritura.

Polakova e Ivenz (2024) desarrollaron un estudio cuasiexperimental con 110 estudiantes universitarios de inglés como lengua extranjera. Los resultados reportaron mejoras en diferentes dimensiones de escritura después de utilizar retroalimentación de ChatGPT, incluyendo concisión, gramática, inclusión de información relevante y uso de voz pasiva.

Tabla 4.
Estudio empírico sobre retroalimentación mediante ChatGPT

Característica	Descripción
Participantes	110 estudiantes universitarios
Área	Inglés como lengua extranjera
Diseño	Cuasiexperimental y mixto
Intervención	Retroalimentación mediante ChatGPT
Producto	Resúmenes escritos
Medidas	Escritura, cuestionarios y grupos focales
Resultados	Mejoras en diferentes dimensiones de escritura
Percepción	Valoración generalmente positiva
Limitación	Contexto específico de escritura en lengua extranjera

El estudio constituye evidencia relevante, aunque su alcance debe interpretarse con prudencia. Los resultados proceden de un contexto específico y no permiten generalizar automáticamente los efectos hacia todas las disciplinas universitarias.

La revisión de Jiao y Wang (2026) amplía esta evidencia al sintetizar 34 estudios desarrollados en 19 países. La revisión concluye que los LLM pueden favorecer productividad, precisión lingüística, participación y alfabetización en retroalimentación,

mientras que la intervención docente mantiene ventajas en dimensiones de escritura de mayor complejidad.

IA generativa frente a retroalimentación humana

Uno de los hallazgos más importantes de la literatura reciente consiste en abandonar una comparación simplista entre “IA” y “docente”.

La evidencia sugiere que ambas formas de retroalimentación poseen fortalezas diferentes.

Tabla 5.
Comparación funcional entre retroalimentación mediante IA y retroalimentación docente

Criterio	IA generativa	Docente
Rapidez	Muy alta	Limitada por disponibilidad
Disponibilidad	Continua	Condicionada por horarios
Personalización superficial	Alta	Alta
Corrección lingüística	Alta	Alta
Revisión iterativa	Muy alta	Variable
Comprensión contextual	Variable	Alta
Interpretación pedagógica	Limitada	Alta

Criterio	IA generativa	Docente
Desarrollo argumentativo	Variable	Alta
Priorización de necesidades	Variable	Alta
Comprensión del contexto institucional	Limitada	Alta
Conocimiento del estudiante	Limitado	Alto cuando existe acompañamiento
Supervisión ética	No autónoma	Profesional
Responsabilidad evaluativa	No debe delegarse	Corresponde al docente

Los resultados de revisiones recientes respaldan esta complementariedad. La revisión de 104 estudios realizada por Lipnevich et al. (2026) encontró variabilidad considerable en los resultados, señalando que los enfoques híbridos que combinan IA y supervisión humana presentan resultados consistentemente favorables frente a modelos exclusivamente automatizados.

Discusión

Los resultados permiten identificar tres transformaciones fundamentales en la conceptualización de la retroalimentación educativa.

La primera corresponde a la transición desde una lógica de retroalimentación escasa y diferida hacia una lógica de retroalimentación abundante e inmediata. La IA generativa puede reducir significativamente el tiempo entre producción y comentario. Esta característica tiene especial valor en actividades donde la revisión constituye parte esencial del aprendizaje.

La segunda transformación consiste en pasar de una retroalimentación predominantemente unidireccional hacia una interacción potencialmente dialógica. El estudiante puede preguntar por qué una respuesta se considera débil, solicitar ejemplos, comparar alternativas y reformular su producción.

La tercera transformación, y posiblemente la más relevante, corresponde al cambio de la pregunta “¿qué comentario recibió el estudiante?” hacia “¿qué hizo el estudiante con ese comentario?”.

Este último aspecto resulta fundamental porque la

retroalimentación solamente se convierte en aprendizaje cuando el estudiante puede comprenderla y utilizarla.

El metaanálisis de Han et al. (2025) proporciona evidencia de que la IA generativa puede producir mejoras globales en resultados de aprendizaje, pero la elevada heterogeneidad del efecto obliga a rechazar una interpretación tecnológica determinista. La tecnología no constituye una intervención homogénea: cambia el resultado según la actividad, la duración, el contexto y la manera en que docentes y estudiantes interactúan con ella.

Una segunda consideración se relaciona con la metacognición. Chen y Cheung (2025) encontraron efectos positivos de la IA generativa en diferentes resultados, pero no identificaron un efecto estadísticamente significativo sobre metacognición. Este resultado resulta particularmente importante porque cuestiona la idea de que proporcionar explicaciones automáticamente conduce necesariamente a estudiantes más autorregulados.

La interpretación más plausible es que la autorregulación requiere algo más que acceso a información. Necesita que el estudiante establezca objetivos, supervise su desempeño, interprete evidencia y tome decisiones. La IA puede apoyar cada una de estas acciones, pero no garantiza que el estudiante las realice.

Este hallazgo coincide con una preocupación creciente en la literatura: la posibilidad de que la IA reduzca determinadas demandas cognitivas si el estudiante delega en el sistema tareas que debería realizar por sí



mismo.

Por ello, una retroalimentación pedagógicamente avanzada debería evitar la respuesta inmediata como primera opción. En su lugar, podría estructurarse mediante niveles progresivos:

Pregunta → pista → explicación → ejemplo → solución, únicamente cuando el estudiante requiera avanzar al siguiente nivel.

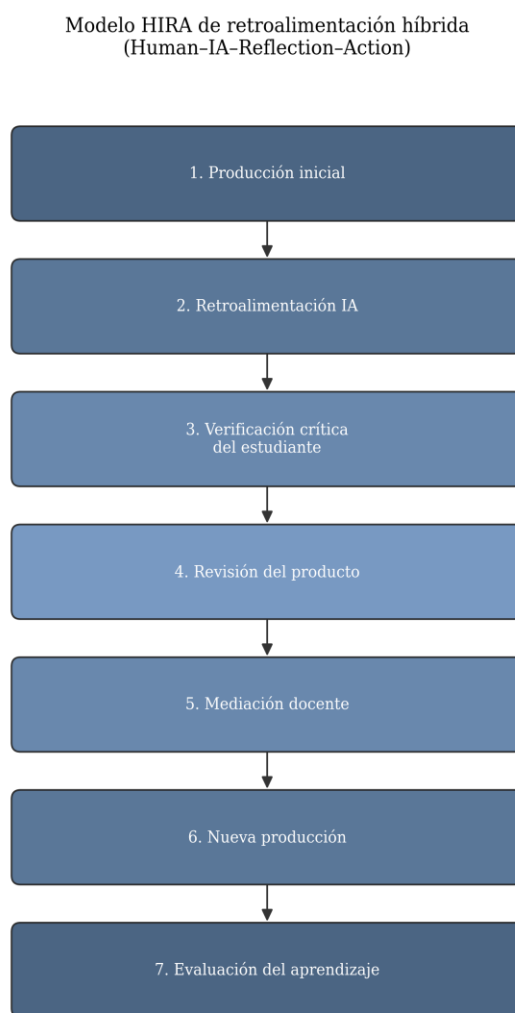
Este modelo permite utilizar la IA como andamiaje sin convertirla automáticamente en sustituto del razonamiento.

Propuesta de modelo de retroalimentación híbrida

A partir de la evidencia analizada se propone un modelo denominado Modelo HIRA: Human–IA–Reflection–Action, orientado a integrar retroalimentación generada por IA con mediación docente.

Figura 2.

Modelo HIRA de retroalimentación híbrida.



El modelo contiene siete fases.

Fase 1. Producción inicial

El estudiante desarrolla la actividad sin delegar completamente la producción en la IA.

Fase 2. Retroalimentación generativa

La herramienta analiza el producto utilizando criterios explícitos.

Fase 3. Verificación

El estudiante determina qué recomendaciones son correctas, cuáles requieren comprobación y cuáles no resultan pertinentes.

Fase 4. Revisión

El estudiante modifica su producción y conserva evidencia de los cambios.

Fase 5. Mediación docente

El docente analiza el proceso y corrige interpretaciones incorrectas.

Fase 6. Nueva producción

El estudiante produce una versión mejorada.

Fase 7. Evaluación

La valoración se centra tanto en el resultado como en el proceso de revisión.

Este modelo presenta una diferencia fundamental respecto del uso convencional de herramientas generativas: la IA interviene dentro del proceso de aprendizaje, pero no reemplaza el proceso de aprendizaje.

Riesgos y desafíos

Alucinaciones

Los sistemas generativos pueden proporcionar información incorrecta con una estructura lingüística

convinciente. En contextos educativos, esto puede generar un problema particularmente complejo porque el estudiante puede interpretar la seguridad de la respuesta como evidencia de su veracidad.

La revisión de Maier et al. (2026) identifica las alucinaciones como uno de los problemas recurrentes de los sistemas de feedback basados en LLM.

Exceso de positividad

Una retroalimentación excesivamente complaciente puede dificultar que el estudiante reconozca debilidades reales de su producción. La utilidad del feedback depende de que exista una relación clara entre evidencia, criterio y recomendación.

Dependencia cognitiva

El estudiante puede desarrollar una conducta de consulta permanente que reduzca la iniciativa para resolver problemas de manera autónoma.

Privacidad

La utilización de plataformas externas puede implicar el tratamiento de información personal y académica. UNESCO recomienda proteger los datos y validar ética y pedagógicamente las herramientas antes de su utilización educativa.

Integridad académica

La frontera entre recibir retroalimentación y delegar la producción intelectual puede volverse difusa. Por ello, las instituciones necesitan diferenciar claramente entre:

- Asistencia.
- Revisión.
- Colaboración.
- Generación.

- Sustitución del trabajo intelectual.

Implicaciones para la educación superior

La evidencia revisada tiene implicaciones específicas para las universidades.

En primer lugar, las instituciones deberían dejar de abordar la IA generativa exclusivamente desde la perspectiva de prohibición o permisividad. El desafío institucional consiste en definir para qué actividades, con qué condiciones y con qué mecanismos de transparencia puede utilizarse.

En segundo lugar, debe fortalecerse la alfabetización en IA. Un estudiante competente no es únicamente aquel que sabe formular prompts, sino aquel que puede evaluar críticamente una respuesta generada, identificar errores, verificar información y justificar las decisiones tomadas.

En tercer lugar, es necesario rediseñar determinadas evaluaciones. Cuando una actividad puede resolverse mediante una única consulta a un modelo de lenguaje, su capacidad para demostrar aprendizaje individual disminuye.

Las evaluaciones deberían incorporar elementos como:

- Defensa oral.
- Bitácoras de aprendizaje.
- Borradores sucesivos.
- Reflexión metacognitiva.
- Justificación de decisiones.
- Resolución de casos.
- Producción contextualizada.
- Revisión de errores.
- Aplicación a problemas nuevos.

En cuarto lugar, el docente debe mantener la responsabilidad sobre la evaluación final. La IA puede asistir en la generación de información, pero no debería convertirse en la autoridad definitiva sobre el aprendizaje de una persona.

Limitaciones del estudio

El presente artículo presenta algunas limitaciones.

En primer lugar, la literatura sobre IA generativa en educación es relativamente reciente y se encuentra en rápida expansión. Esto significa que determinados resultados pueden cambiar conforme aumente la disponibilidad de estudios longitudinales.

En segundo lugar, existe una concentración considerable de investigaciones en educación superior, particularmente en escritura y aprendizaje de idiomas. La revisión de Jiao y Wang (2026), por ejemplo, identifica una fuerte concentración de estudios en contextos universitarios de escritura.

En tercer lugar, existe una importante heterogeneidad metodológica. Los estudios difieren en población, herramienta, duración, criterios de evaluación y condiciones de intervención.

Finalmente, la mayoría de las investigaciones disponibles analiza intervenciones relativamente cortas. Por tanto, todavía existe incertidumbre respecto de los efectos sostenidos de la utilización frecuente de IA sobre autonomía, pensamiento crítico y autorregulación.

Estas limitaciones justifican la necesidad de investigaciones longitudinales y diseños experimentales con comparadores claramente definidos.

Agenda para futuras investigaciones

La evidencia revisada permite identificar cinco líneas prioritarias.

Primera: desarrollar estudios longitudinales que permitan analizar si las mejoras obtenidas mediante IA se mantienen en el tiempo.

Segunda: comparar sistemáticamente tres condiciones: retroalimentación humana, retroalimentación exclusivamente mediante IA y retroalimentación híbrida.

Tercera: analizar no solamente el producto final, sino también los procesos cognitivos utilizados por el estudiante para interpretar la retroalimentación.

Cuarta: ampliar la investigación hacia disciplinas diferentes de idiomas y escritura, especialmente ciencias, ingeniería, salud, economía y educación.

Quinta: incorporar variables de equidad, privacidad, alfabetización digital y dependencia tecnológica.

Estas líneas permitirían avanzar desde una etapa caracterizada por estudios de adopción y percepción hacia una ciencia de la retroalimentación asistida por IA fundamentada en evidencia causal y longitudinal.

CONCLUSIONES

La Inteligencia Artificial generativa constituye una herramienta con potencial significativo para transformar la retroalimentación del aprendizaje. La evidencia reciente muestra resultados favorables sobre diferentes dimensiones educativas, particularmente cuando la herramienta permite revisión iterativa, explicación de errores y apoyo personalizado.

Sin embargo, los resultados no respaldan una interpretación determinista según la cual la

incorporación de IA generativa produciría automáticamente mejores aprendizajes. El metaanálisis de Han et al. (2025) demuestra precisamente esta complejidad: aunque se identifica un efecto positivo global, la heterogeneidad entre estudios es muy elevada.

La principal fortaleza de la IAGen se encuentra en su capacidad para proporcionar retroalimentación rápida, disponible y repetible. Esta característica puede reducir determinadas limitaciones temporales de los procesos educativos y facilitar ciclos más frecuentes de revisión.

No obstante, la evidencia reciente indica que las ventajas de la IA disminuyen cuando la retroalimentación exige interpretación contextual, desarrollo argumentativo, priorización de problemas o acompañamiento metacognitivo. En estos ámbitos, la mediación docente conserva un papel central.

Por tanto, el modelo educativo más consistente con la evidencia disponible no es la sustitución del docente mediante inteligencia artificial, sino la complementariedad entre inteligencia artificial, estudiante y docente.

La IA puede proporcionar información; el estudiante debe interpretarla; el docente debe contextualizarla y validarla.

Desde esta perspectiva, la retroalimentación asistida por IA adquiere verdadero valor cuando se convierte en un proceso de interacción que conduce a la reflexión, revisión y mejora. El objetivo no debería ser generar más comentarios, sino conseguir que cada comentario produzca una acción cognitiva significativa.

En consecuencia, la IAGen debe ser entendida como un instrumento de andamiaje para la retroalimentación, integrado en diseños pedagógicos

que preserven la autonomía, la responsabilidad intelectual, la privacidad y la capacidad crítica del estudiante.

La evidencia disponible permite sostener que el futuro de la retroalimentación educativa no se encuentra en elegir entre inteligencia artificial o docentes, sino en diseñar sistemas pedagógicos capaces de aprovechar las fortalezas de ambas formas de inteligencia bajo criterios de responsabilidad, transparencia y finalidad educativa.

REFERENCIAS BIBLIOGRÁFICAS

- Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis. *Educational Research Review*, 48, 100737. <https://doi.org/10.1016/j.edurev.2025.100737>
- Han, X., Peng, H., & Liu, M. (2025). The impact of GenAI on learning outcomes: A systematic review and meta-analysis of experimental studies. *Educational Research Review*, 48, 100714. <https://doi.org/10.1016/j.edurev.2025.100714>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Jiao, Y., & Wang, Q. (2026). Large language models for formative feedback in writing instruction: A systematic review of classroom interventions, feedback quality, and student outcomes. *Frontiers in Education*, 11, 1834085. <https://doi.org/10.3389/educ.2026.1834085>
- Lipnevich, A. A., Taranto, C. A., DeLuca, C., Nukpetsi, E., Rickey, N., & Hopfenbeck, T. (2026). Mapping the landscape of AI-driven feedback in education: A scoping review. *Frontiers in Education*, 11, 1799346. <https://doi.org/10.3389/educ.2026.1799346>
- Maier, U., Seibold, M., & Klotz, C. (2026). LLM-generated formative feedback in education: A qualitative systematic literature review. *Computers and Education: Open*, 10, 100374. <https://doi.org/10.1016/j.caeo.2026.100374>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
- UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.
- UNESCO. (2023). Guía para el uso de IA generativa en educación e investigación. UNESCO.