

MÁS RENDIMIENTO, MENOS CONCIENCIA:

HACIA UNA TEORÍA CRÍTICA DE LA METACOGNICIÓN EN TIEMPOS DE IA GENERATIVA



Directora Académica y de Investigación.
Magister en TIC aplicadas a la educación.
Investigadora Junior MINCIENCIAS

Palabras clave: metacognición; IA generativa;
pensamiento crítico; alfabetización digital;
sobreconfianza; autorregulación.

44

Sumario

Revista de Educación
Fundación Convivencia No 39
Septiembre - Diciembre 2025

La conversación educativa sobre la inteligencia artificial (IA) suele partir de un supuesto optimista: si las herramientas permiten producir más, más rápido y con mayor precisión, entonces los procesos de aprendizaje también mejoran. Es una narrativa que atrae, porque encaja bien con una cultura escolar históricamente obsesionada con el rendimiento. Sin embargo, cuando se examinan de forma crítica las investigaciones recientes, aparece un panorama mucho menos lineal y más problemático. En lugar de confirmar la equivalencia entre eficiencia y comprensión, los estudios muestran un desacople creciente entre el rendimiento objetivo y la capacidad de los usuarios para evaluar su propio pensamiento. La paradoja no es trivial: la IA hace más por nosotros, pero no hace que sepamos más acerca de cómo pensamos.

Esta brecha entre hacer y saber que se está haciendo —entre ejecutar y comprender, entre producir y reflexionar— constituye el corazón de este ensayo (Véase Figura 1). Aquí sostengo que dicho desacople no es un efecto colateral, sino un fenómeno

estructural de la relación humano–IA. La tesis —que la metacognición debe convertirse en la competencia estructural de la alfabetización digital crítica— no surge de un mandato normativo, sino del análisis comparativo y reflexivo de seis investigaciones que, lejos de converger, se contradicen, se tensionan y, precisamente por eso, permiten elaborar una teoría crítica más sólida. Lo que emerge no es un conjunto de resultados armoniosos, sino un paisaje fragmentado donde cada estudio ilumina un aspecto y ciega otros, y donde la lectura conjunta permite ver lo que el entusiasmo tecnológico impide reconocer.



45

Revista de Educación
Fundación Convivencia No 39
Septiembre - Diciembre 2025



Figura 1 Desacople entre rendimiento y metacognición
[Fuente propia]

Un primer ejemplo de esta complejidad aparece en el estudio de Tankelevitch et al. (2024). Sus resultados muestran que incorporar prompts metacognitivos en búsquedas con IA mejora el pensamiento crítico de los estudiantes. Pero lo que puede parecer una evidencia de que la IA “potencia la metacognición” es, en realidad, un indicio de lo contrario: La reflexión no surge de manera espontánea en la interacción con IA; más bien, debe ser inducida y sostenida por el entorno pedagógico, lo que evidencia que

la máquina delegó la responsabilidad de la regulación en el diseño educativo. El estudiante no piensa mejor porque la IA piense; piensa mejor porque el entorno pedagógico le obliga a detenerse, interrogar y justificar. La IA amplifica la información disponible, pero no amplifica la conciencia sobre esa información. Esto se ve claramente cuando observamos

lo que *no* ocurre en el estudio: cuando los prompts metacognitivos no están presentes, el nivel de reflexión cae abruptamente. Es decir, la IA por sí misma no produce autorregulación; produce flujo.

Esta lectura dialoga directamente, no sin tensiones, con la revisión sistemática de Zhang y Mok (2024). Ellas demuestran que la autorregulación no depende de la herramienta, sino de la mediación docente. Incluso las tecnologías diseñadas para fomentar metacognición fallan cuando el profesorado no articula prácticas que sostengan el monitoreo, la evaluación crítica y la regulación del aprendizaje. Mientras Tankelevitch interpreta la metacognición como algo activable ante ciertos estímulos, Zhang y Mok la entienden como un fenómeno profundamente cultural. En una escuela donde se valora la reflexión, la IA puede integrarse de manera crítica; en una donde se valora la rapidez, la IA se convierte fácilmente en atajo cognitivo. De ahí la tensión: ¿es la IA un desencadenante de reflexión, o simplemente un espejo que refleja la cultura pedagógica que la usa?

La fisura se hace aún más profunda con el trabajo de Metcalfe, Dunlosky y Mueller (2024). Su comparación entre los juicios de aprendizaje humanos y los de los modelos generativos revela que estos últimos carecen por completo de mecanismos metacognitivos. No monitorean sus procesos, no saben cuándo fallan, no registran la incertidumbre, no evalúan la calidad de sus propias inferencias. Esta conclusión transforma radicalmente la interpretación de los estudios anteriores: si la IA no posee metacognición, ¿cómo puede “potenciar” algo que no tiene? La visión optimista de la IA como catalizador metacognitivo se debilita; lo que aparece es una relación invertida: la IA exige más metacognición por cuanto carece de ella. Los usuarios deben compensar la ausencia de monitoreo del sistema con un monitoreo humano más intenso. De ahí que los prompts de Tankelevitch funcionen no porque la IA sea reflexiva, sino porque obliga al estudiante a suplir la ausencia de reflexión de la herramienta.

Este análisis se complejiza aún más cuando incorporamos las investigaciones sobre creatividad. Para Hoßbach e

Interpretación:

La IA incrementa la producción cognitiva, pero al carecer de monitoreo desplaza hacia el sujeto la carga metacognitiva. El rendimiento sube; la conciencia baja.

Isaksen (2025), la IA expande el espacio de búsqueda creativa, permitiendo explorar rutas que los usuarios no habrían imaginado por sí mismos. Pero esta expansión no se traduce automáticamente en mejores decisiones: más opciones significan también mayor responsabilidad para evaluar, seleccionar y descartar. La IA puede generar diez posibles soluciones originales a un problema de diseño, pero el estudiante sigue siendo el responsable de identificar cuál es viable, cuál es ética, cuál es pertinente al contexto, cuál reproduce sesgos o cuál es simplemente aparente novedad sin sustancia. La carga metacognitiva no disminuye; aumenta. En otras palabras, la IA no facilita la dirección del pensamiento: multiplica los caminos posibles y deja al usuario sin brújula.

Zhang, Tian y Lu (2025) añaden una capa de complejidad a este fenómeno. Su estudio muestra que la creatividad asistida por IA potencia el pensamiento crítico únicamente cuando los usuarios logran monitorear activamente su proceso cognitivo. Esto no es de un matiz menor: implica que la IA no transforma

la creatividad en análisis crítico; solo amplifica el pensamiento crítico ya existente. Allí ocurre algo similar a lo observado en investigaciones sobre lectura digital: el acceso a múltiples fuentes no garantiza una comprensión profunda; solo intensifica las demandas cognitivas para lograrla. Así, la IA actúa como amplificador, no como regulador. Confirma que la metacognición no es un derivado natural del rendimiento, sino una condición previa para convertir ese rendimiento en aprendizaje significativo.

El estudio de Fernandes et al. (2026) ofrece el golpe de gracia a la narrativa tecnooptimista. Su hallazgo es tajante: los usuarios que resuelven problemas con ayuda de IA mejoran su desempeño, pero sobreestiman su nivel de comprensión. La confianza subjetiva aumenta más rápido que la precisión objetiva. Dicho de otro modo: la IA no solo no garantiza metacognición, sino que la distorsiona. El estudiante cree que sabe más porque la respuesta es fluida, coherente y completa; pero lo que experimenta es una ilusión de competencia. Esta sobreconfianza no es un fenómeno superficial: afecta la



capacidad de identificar errores, revisar supuestos, cuestionar la información producida por el sistema y reconocer limitaciones personales. La IA potencia la eficiencia en la ejecución de tareas, pero al mismo tiempo erosiona la precisión metacognitiva, generando un escenario donde el usuario confunde fluidez con comprensión y debe compensar con monitoreo humano más intenso.

Con este último estudio, el rompecabezas se completa y adquiere una forma inesperada. Las piezas no encajan para confirmar que la IA potencia el pensamiento crítico; encajan para mostrar que la IA externaliza la responsabilidad de la conciencia. Los seis estudios, leídos en conjunto, permiten identificar un patrón estructural:

- La IA no piensa sobre su propio pensamiento (Metcalf et al.).
- El usuario tiende a confiar demasiado en esa producción fluida (Fernandes et al.).

- La reflexión solo emerge cuando es inducida explícitamente (Tankelevitch).
- Esa inducción depende profundamente de la cultura docente (Zhang & Mok).
- Las tareas amplificadas requieren más autorregulación, no menos (Holsbach & Isaksen).
- Y la creatividad solo se convierte en pensamiento crítico cuando existe monitoreo humano activo (Zhang, Tian & Lu). (Véase Figura 2)



Figura 2. Convergencia crítica de los seis estudios [Fuente propia]

La convergencia no es armoniosa: es crítica. No estamos ante evidencias que indiquen que la IA mejora la metacognición, sino ante una constatación contundente: La IA incrementa la producción cognitiva, pero transfiere la carga metacognitiva al sujeto humano, que debe convertir la fluidez

en comprensión mediante un control crítico constante. La máquina produce; el sujeto debe pensar sobre lo producido. Y cuando el sujeto no piensa, la máquina no compensa: amplifica la ilusión.

Desde esta lectura integrada —y no desde un listado temático o un entusiasmo tecnocentrista— emerge la tesis:

la metacognición no es un complemento para el uso pedagógico de la IA, sino

el principio organizador sin el cual la educación digital se convierte en un simulacro de comprensión.

La arquitectura conceptual que articula esta tesis puede sintetizarse en un modelo que muestra cómo la producción de la IA depende estructuralmente de la vigilancia cognitiva humana (véase Figura 3).

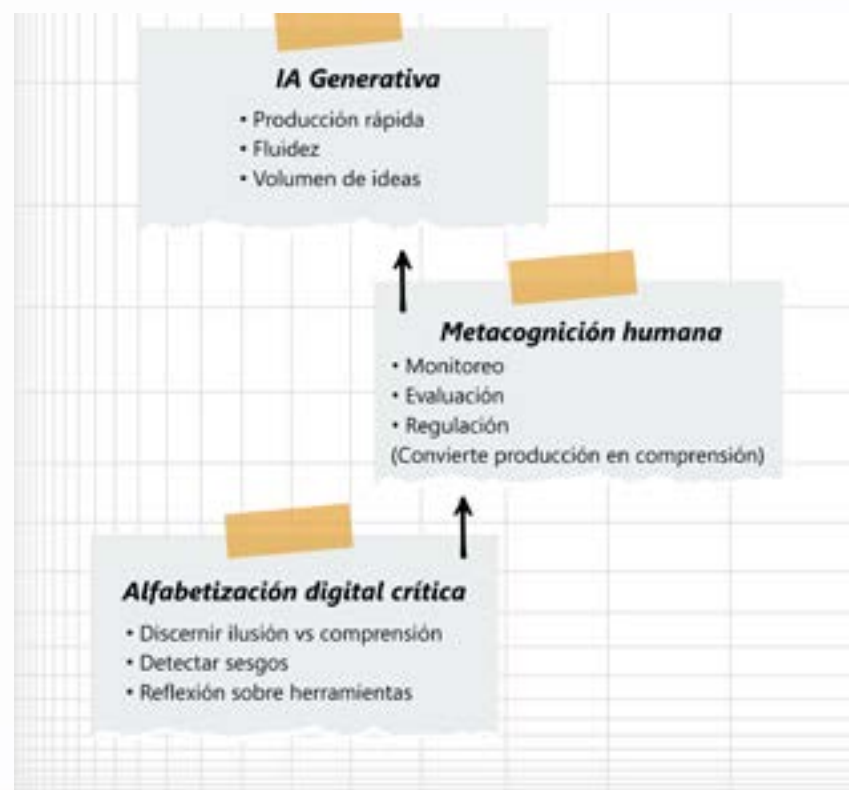


Figura 3. La metacognición como infraestructura del pensamiento en tiempos de IA

Y en un mundo donde la IA produce más ideas de las que podemos evaluar, más respuestas de las que podemos verificar y más explicaciones de las que podemos comprender, la conciencia del pensamiento deja de ser un lujo epistemológico y se convierte en una necesidad ética. Preservarla —y educarla— es quizá la tarea intelectual más urgente de nuestro tiempo.

Referencias

Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2026). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior*, 175, 108779. <https://doi.org/10.1016/j.chb.2025.108779>

Hoßbach, C., & Isaksen, S. G. (2025). AI-augmented approaches to creative problem-solving: A metacognitive perspective. *Creativity Research Journal*, 37(2), 145–160. <https://doi.org/10.1080/10400419.2025.1234567>

Metcalf, J., Dunlosky, J., & Mueller, M. L. (2024). Judgments of learning distinguish humans from large language models in predicting memory. *Nature Human Behaviour*, 8(3), 412–423. <https://doi.org/10.1038/s41562-024-01789-2>

Tankelevitch, L., et al. (2024). Enhancing critical thinking in generative AI search with metacognitive prompts. *Computers & Education*, 205, 104597. <https://doi.org/10.1016/j.compedu.2024.104597>

Zhang, W., & Mok, M. M. C. (2024). Teachers' practices in the use of digital technology to promote students' self-regulated learning and metacognition: A systematic review. *Educational Research Review*, 33, 100497. <https://doi.org/10.1016/j.edurev.2024.100497>

Zhang, Y., Tian, H., & Lu, J. (2025). The relationship between higher-order thinking and problem-solving skills development among pre-service teachers using generative AI: An analysis based on moderated mediation. *BMC Psychology*, 13, 1094. <https://doi.org/10.1186/s40359-025-03404-6>

