

# Retroalimentación formativa mediada por inteligencia artificial y conciencia metacognitiva: un modelo teórico para las pedagogías activas AI-Driven Formative Feedback and Metacognitive Awareness: A Theoretical Model for Active Pedagogies.

Reinaldo Antonio Guerrero Chirinos<sup>1</sup>, María de los Ángeles Guamán Coronel<sup>2</sup>, Nancy Evangelista Ricardo Domínguez<sup>3</sup>, Fabián Alejandro Banderas Benítez<sup>4</sup>, Manuel Enrique Escanio Cortes<sup>5</sup>, María de Jesús Mejía Díaz<sup>6</sup>, Yamira Iraisá Herrera García<sup>7</sup>

<sup>1</sup> Universidad Técnica Particular de Loja (UTPL), Loja, Ecuador.  
(raguerrero12@utpl.edu.ec)  
ORCID: 0000-0003-0499-7453

<sup>2</sup> Universidad Técnica Particular de Loja (UTPL), Loja, Ecuador.  
(mguaman1@utpl.edu.ec)  
ORCID: 0000-0003-1778-8044

<sup>3</sup> Universidad Católica de Santiago de Guayaquil, Ecuador.  
(nancy.ricardo@cu.ucsg.edu.ec)  
ORCID: 0000-0002-0251-1862

<sup>4</sup> Escuela Superior Politécnica de Chimborazo (ESPOCH), Ecuador.  
(fabian.banderasb@epoch.edu.ec)  
ORCID: 0000-0001-6242-873X

<sup>5</sup> Universidad Nacional Autónoma de México (UNAM), Mexico.  
(manuel.escanio@imss.gob.mx)  
ORCID: 0000-0003-2946-9315

<sup>6</sup> Instituto Suizo Universitario (ISU), Switzerland; Universidad Nacional Autónoma de México (UNAM), Mexico.  
(dramarimejia@gmail.com)  
ORCID: 0009-0001-1186-8660

<sup>7</sup> Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Peru.  
(Yamira.herrera@untrm.edu.pe)  
ORCID: 0000-0003-2881-5992

**Abstract:** The integration of artificial intelligence (AI) systems in educational environments raises fundamental questions about how technologically mediated formative feedback can enhance students' metacognitive awareness. This theoretical-conceptual article proposes the METACOGNITO model, a name that evokes the

articulation between metacognition and AI-mediated cognition. The model integrates four theoretical domains: (1) theories of metacognition and self-regulated learning; (2) the principles and mechanisms of AI-mediated formative feedback, including the literature on metacognitive prompts; (3) contemporary frameworks of cognitive engagement and active student-centered methodologies; and (4) AI ethics in education. Through a structured narrative review with framework synthesis, the model describes a five-phase processual cycle (activation, dynamic feedback, metacognitive reflection, regulatory adjustment, and transfer) in which AI agents act as dynamic scaffolds that amplify the learner's internal monitoring and cognitive regulation processes. The original contribution lies in the metacognitive connector construct, a design component distinguished from classical metacognitive prompts by its coupling with the feedback cycle, its multilevel intentionality, and its adaptivity. The model is accompanied by a conceptual map and five testable propositions for future empirical research. Implications for instructional design, teacher training, and educational policy are discussed, emphasizing the need for an ethical, inclusive, and culturally sensitive approach to AI implementation in active pedagogies.

**Keywords:** formative feedback; metacognitive awareness; artificial intelligence in education; active methodologies; self-regulated learning; theoretical model; metacognitive connector

## Introduction

El campo de la educación superior atraviesa una transformación paradigmática impulsada por la convergencia de tres fuerzas: la creciente adopción de metodologías activas de enseñanza-aprendizaje, el auge de la inteligencia artificial generativa como herramienta pedagógica, y el renovado interés científico por los procesos metacognitivos como predictores del aprendizaje profundo y duradero. Esta confluencia genera un espacio de intersección teórica con enorme potencial para reconfigurar la pedagogía contemporánea.

La retroalimentación formativa, entendida como el proceso continuo de información que orienta al estudiante hacia la mejora de su desempeño, ha sido reconocida durante décadas como uno de los factores de mayor impacto en los logros de aprendizaje (Hattie & Timperley, 2007). Sin embargo, su implementación en contextos de educación a gran escala ha enfrentado barreras estructurales significativas: la carga de trabajo docente, el número de estudiantes por aula y la dificultad de ofrecer retroalimentación oportuna y personalizada de manera sostenible (Banihashem et al., 2025). La aparición de sistemas de IA, especialmente los modelos de lenguaje de gran escala (LLM) y los sistemas de tutoría inteligente (ITS), ofrece por primera vez la posibilidad técnica de escalar la retroalimentación formativa sin sacrificar su calidad ni su oportunidad (Lan & Zhou, 2025).

Paralelamente, la investigación en metacognición ha demostrado de manera consistente que los estudiantes que desarrollan conciencia y regulación de sus propios procesos cognitivos obtienen mejores resultados académicos, mayor autonomía como aprendices y mayor capacidad de transferencia del conocimiento (Järvelä & Bannert, 2021; Tsakeni et al., 2025). No obstante, la metacognición no emerge espontáneamente: requiere entornos que la promuevan activamente, que ofrezcan oportunidades para el monitoreo, la evaluación y el ajuste estratégico del propio aprendizaje.

Las metodologías activas, que incluyen el aprendizaje basado en problemas (ABP), el aprendizaje basado en proyectos (PjBL), el aula invertida y el aprendizaje colaborativo, entre otras, constituyen precisamente ese tipo de entorno. Al colocar al estudiante como agente constructor de su propio conocimiento, estas metodologías crean condiciones estructurales para que la metacognición emerja y se consolide. Sin embargo, la sola aplicación de metodologías activas no garantiza el desarrollo metacognitivo; se requieren andamiajes específicos que activen y refuercen los procesos de reflexión, monitoreo y regulación (Chi & Wylie, 2014; Bannert, 2009).

### 1.1. Brecha de investigación y pregunta orientadora

Pese a los avances independientes en cada uno de los tres dominios mencionados, la investigación contemporánea presenta una brecha significativa en su articulación. Por una parte, la literatura sobre metacognición y aprendizaje autorregulado ha desarrollado modelos cíclicos sólidos (Järvelä & Bannert, 2021; Molenaar et al., 2023) pero ha incorporado solo recientemente, y de manera fragmentaria, el papel de la IA como agente regulador externo (Järvelä et al., 2023; Lan & Zhou, 2025). Por otra parte, los estudios sobre retroalimentación generada por IA han documentado su efectividad pedagógica (Bellhäuser et al., 2023; Banihashem et al., 2025) pero rara vez integran explícitamente las

fases del ciclo metacognitivo en el diseño del sistema de retroalimentación. Finalmente, la literatura consolidada sobre prompts metacognitivos (Bannert, 2009; Guo, 2022; Wong et al., 2021) ha demostrado el efecto positivo de las preguntas reflexivas estratégicamente diseñadas, pero no ha sido integrada con los nuevos sistemas de retroalimentación generativa con IA ni con los marcos de engagement cognitivo en metodologías activas (Chi & Wylie, 2014; Fredricks et al., 2004).

En consecuencia, no existe en la literatura un modelo teórico que articule procesualmente los tres dominios anteriores y que sitúe la retroalimentación formativa con IA como mecanismo específico de andamiaje metacognitivo dentro de entornos pedagógicos activos. La pregunta de investigación que orienta el presente trabajo es, por tanto, la siguiente: ¿cómo puede la retroalimentación formativa mediada por inteligencia artificial andamiar de manera sistemática el desarrollo de la conciencia metacognitiva de los estudiantes en entornos de metodologías activas?

El artículo se estructura siguiendo la organización IMRyD adaptada a su naturaleza teórica. La sección 2 desarrolla el marco teórico que sustenta los cuatro dominios conceptuales del modelo. La sección 3 describe la metodología de revisión narrativa estructurada y de síntesis basada en marcos. La sección 4 presenta los resultados, esto es, el modelo METACOGNITO con sus componentes, fases, proposiciones contrastables y mapa conceptual. La sección 5 discute los aportes, las tensiones de diseño y las implicaciones del modelo. La sección 6 cierra con conclusiones, limitaciones y futuras líneas de investigación.

## Marco teórico

### 2.1. Metacognición y conciencia metacognitiva: fundamentos conceptuales

La metacognición, definida clásicamente como el conocimiento y la cognición acerca de los fenómenos cognitivos (Flavell, 1979), ha evolucionado conceptualmente hasta convertirse en un constructo multidimensional que engloba tanto el conocimiento metacognitivo (el saber sobre la propia cognición) como la regulación metacognitiva (el control activo de los propios procesos cognitivos durante el aprendizaje) (Azevedo et al., 2022; Järvelä & Bannert, 2021).

La conciencia metacognitiva, como dimensión específica dentro de este constructo, hace referencia a la capacidad del aprendiz para reconocer, monitorear y evaluar sus propios estados cognitivos en tiempo real: saber cuándo comprende y cuándo no, identificar qué estrategias son más efectivas para una tarea dada, y ajustar su comportamiento de aprendizaje en consecuencia (Wang et al., 2022). Esta capacidad no es estática ni unidimensional; se desarrolla de manera progresiva a través de la interacción con entornos que demandan reflexión explícita sobre el proceso de aprendizaje (Tsakeni et al., 2025).

El marco del aprendizaje autorregulado (SRL) proporciona la arquitectura teórica más comprehensiva para entender cómo la metacognición se integra en el proceso de aprendizaje. El modelo de Zimmerman (2000), ampliamente citado y revisado en investigaciones contemporáneas sobre SRL (Järvelä & Bannert, 2021; Molenaar et al., 2023; Fan et al., 2023), describe tres fases interdependientes: la previsión (forethought), donde el aprendiz establece metas y planifica estrategias; la ejecución (performance), donde aplica y monitorea sus estrategias; y la autorreflexión (self-reflection), donde evalúa sus resultados y realiza atribuciones causales. Investigaciones recientes han ampliado este modelo para incluir las formas en que la IA puede actuar como coregulador externo del proceso de aprendizaje (Lan & Zhou, 2025; Järvelä et al., 2023).

Particularmente relevante para este trabajo es la distinción entre la metacognición individual y las formas colectivas de regulación del aprendizaje. En entornos de metodologías activas, donde el aprendizaje colaborativo es frecuente, emerge el constructo de regulación compartida socialmente (SSRL), que describe los procesos mediante los cuales grupos de aprendices comonitorizan y coregulan sus procesos de aprendizaje conjunto (Järvelä et al., 2023; Edwards et al., 2025). La integración de la IA en estos contextos añade una dimensión adicional: la regulación humano-IA compartida, donde los sistemas de IA participan activamente como agentes de regulación externa que complementan los

procesos metacognitivos humanos (Järvelä et al., 2023).

Investigaciones recientes han advertido, no obstante, sobre el riesgo de la pereza metacognitiva (metacognitive laziness), un fenómeno por el cual la facilidad de acceso a retroalimentación generada por IA puede paradójicamente inhibir el esfuerzo metacognitivo del estudiante, reduciéndolo a un consumidor pasivo de respuestas en lugar de un agente reflexivo de su propio aprendizaje (Fan et al., 2025; Zhang et al., 2024). Este riesgo, asociado al fenómeno más amplio de la descarga cognitiva (cognitive offloading) hacia los sistemas de IA, debe incorporarse en cualquier modelo teórico que articule IA y metacognición.

## 2.2. Retroalimentación formativa mediada por IA: principios y mecanismos

La retroalimentación formativa constituye uno de los constructos más robustamente respaldados por la evidencia empírica en la investigación educativa. El modelo de Hattie & Timperley (2007), que sigue siendo el marco de referencia más citado en la literatura contemporánea, distingue cuatro niveles de retroalimentación: sobre la tarea, sobre el proceso, sobre la autorregulación y sobre el yo (self). Los niveles de proceso y autorregulación son los más potentes para el aprendizaje profundo, al dirigir la atención del estudiante hacia los mecanismos de su propia cognición (Banihashem et al., 2025).

Nicol & Macfarlane-Dick (2006) complementaron este marco con sus siete principios de buenas prácticas de retroalimentación, entre los cuales destacan: la clarificación de los estándares de desempeño, el fomento de la autoevaluación, la provisión de información de calidad sobre el progreso, y el estímulo al diálogo reflexivo entre docente y estudiante. Estos principios constituyen una agenda de diseño que los sistemas de IA actuales están comenzando a operacionalizar de manera escalable.

La emergencia de los modelos de lenguaje de gran escala (LLM) como ChatGPT ha transformado radicalmente las posibilidades de la retroalimentación automatizada. A diferencia de los sistemas de retroalimentación basados en reglas de generaciones anteriores, los LLM pueden generar retroalimentación contextualizada, elaborada y adaptativa que va más allá de la simple corrección de errores (Bellhäuser et al., 2023; Urban et al., 2025). Los sistemas de tutoría inteligente (ITS) de nueva generación, como MetaTutor, integran estas capacidades con modelos del aprendiz dinámicos que adaptan la retroalimentación al estado cognitivo y metacognitivo específico del estudiante en cada momento (Azevedo et al., 2022).

Sin embargo, la investigación reciente también ha identificado limitaciones significativas en la retroalimentación generada por IA. Urban et al. (2025) encontró que las creencias epistemológicas y la precisión metacognitiva de los estudiantes modulan de manera significativa la calidad de la integración de la retroalimentación generada por IA en el trabajo académico. Asimismo, la retroalimentación de la IA tiende a ser más efectiva cuando es procesada activamente por el estudiante, generando reflexión metacognitiva, que cuando es simplemente aceptada de manera acrítica (Fan et al., 2025; Xu et al., 2025).

La retroalimentación mediada por IA puede operar en distintas temporalidades: inmediata (durante la tarea), diferida (al finalizar una actividad) y prospectiva (orientada a tareas futuras). Cada temporalidad activa diferentes procesos metacognitivos. La retroalimentación inmediata favorece el monitoreo en tiempo real; la diferida potencia la reflexión evaluativa; y la prospectiva activa la planificación estratégica anticipatoria. Un modelo teórico comprehensivo debe articular estas tres temporalidades con las fases del ciclo metacognitivo (Molenaar et al., 2023; Lim et al., 2023).

La alfabetización en retroalimentación (feedback literacy) emerge como un constructo mediador crítico entre la retroalimentación formativa y sus efectos metacognitivos. Carless & Winstone (2023) definen la alfabetización docente en retroalimentación como el conjunto de competencias para diseñar, implementar y gestionar procesos de retroalimentación que promuevan la agencia del estudiante; mientras que la alfabetización estudiantil en retroalimentación implica la capacidad de interpretar, valorar y utilizar la retroalimentación para regular el propio aprendizaje (Lipnevich & Smith, 2022). En el contexto de la IA, esta alfabetización debe extenderse para incluir la capacidad de evaluar críticamente la retroalimentación generada por sistemas automatizados.

## 2.3. Andamiajes metacognitivos: prompts y más allá

Una línea de investigación particularmente relevante para el modelo aquí propuesto es la

dedicada a los prompts metacognitivos. Bannert (2009) sentó las bases conceptuales de este campo al definirlos como pistas, claves o preguntas dirigidas a la metacognición del aprendiz, capaces de activar conocimientos y estrategias que el estudiante posee pero no aplica espontáneamente. La revisión meta-analítica de Guo (2022) sobre estudios experimentales en entornos de aprendizaje computarizado demostró que los prompts metacognitivos producen efectos positivos significativos sobre las actividades de SRL ( $g = 0,50$ ) y sobre los resultados de aprendizaje, con tamaños de efecto moderados.

Wong et al. (2021), en estudios sobre cursos en línea masivos, mostraron que los prompts orientados a la planificación, el monitoreo y la evaluación pueden incrementar la autorregulación de los estudiantes cuando se diseñan de manera adaptativa al perfil del aprendiz. No obstante, esta literatura presenta tres limitaciones que el modelo METACOGNITO busca subsanar. En primer lugar, los prompts metacognitivos clásicos suelen operar de manera independiente del ciclo de retroalimentación, esto es, se entregan como intervenciones aisladas en momentos específicos y no como parte de un flujo integrado de información formativa. En segundo lugar, la literatura clásica de prompts trabaja preferentemente con sistemas basados en reglas predefinidas, sin aprovechar la capacidad generativa y adaptativa de los LLM contemporáneos. En tercer lugar, la mayoría de estudios sobre prompts metacognitivos no se ha articulado con marcos de engagement cognitivo, dejando sin abordar la cuestión de cómo el prompt se integra en el ecosistema pedagógico activo más amplio.

El conector metacognitivo, propuesto en este artículo como aporte central del modelo METACOGNITO, hereda la lógica de los prompts metacognitivos pero se distingue de ellos en tres dimensiones específicas que se desarrollan en la sección 4 de Resultados: (a) su acoplamiento estructural a cada pieza de retroalimentación formativa; (b) su intencionalidad multinivel, que articula los cuatro niveles de Hattie & Timperley (2007) con las dimensiones del SRL de Zimmerman (2000); y (c) su adaptatividad dinámica al perfil metacognitivo cambiante del estudiante a través del modelo del aprendiz mantenido por el sistema de IA.

#### 2.4. Engagement cognitivo y metodologías activas: el marco ICAP

Las metodologías activas representan una familia de enfoques pedagógicos que comparten el principio de posicionar al estudiante como agente activo en la construcción de su conocimiento. Entre las más consolidadas por la evidencia empírica se encuentran el aprendizaje basado en problemas (ABP), el aprendizaje basado en proyectos (PjBL), el aula invertida (flipped classroom) y el aprendizaje colaborativo. Todas ellas crean condiciones estructurales que demandan, y por tanto desarrollan, procesos metacognitivos (Chiu et al., 2023).

Para fundamentar teóricamente la noción de "actividad" en estas metodologías, el modelo METACOGNITO recurre al marco ICAP de Chi & Wylie (2014). Este marco distingue cuatro modos de engagement cognitivo del estudiante con los materiales de aprendizaje: pasivo (recepción sin elaboración), activo (manipulación física o atención sostenida), constructivo (generación de nuevas inferencias o conocimientos) e interactivo (construcción colaborativa con otros agentes). La hipótesis ICAP predice que los resultados de aprendizaje aumentan en la dirección Pasivo < Activo < Constructivo < Interactivo. Esta jerarquía resulta crítica para el modelo aquí propuesto porque sugiere que la efectividad de la retroalimentación con IA depende del modo de engagement que el diseño instruccional active: una retroalimentación entregada en un contexto pasivo producirá efectos cualitativamente distintos a la misma retroalimentación entregada en un contexto interactivo.

Complementariamente, la teoría del engagement de Fredricks et al. (2004) distingue tres dimensiones del compromiso del estudiante con el aprendizaje: comportamental, emocional y cognitiva. El modelo METACOGNITO se concentra en la dimensión cognitiva del engagement (la inversión psicológica del estudiante en la comprensión y el dominio del contenido), pero reconoce que esta dimensión está modulada por las otras dos. Reeve (2013) introdujo posteriormente el constructo de engagement agéntico, que se refiere a la iniciativa proactiva del estudiante para enriquecer su propio entorno de aprendizaje, y que conecta directamente con el principio de agencia estudiantil que el modelo METACOGNITO incorpora como característica del entorno pedagógico activo.

El aprendizaje basado en problemas, en particular, crea situaciones de incertidumbre cognitiva que obligan al estudiante a monitorizar continuamente su comprensión, identificar brechas de conocimiento y ajustar sus estrategias de resolución. Esta estructura es isomórfica con el ciclo metacognitivo descrito por Flavell: conciencia del problema, planificación, monitoreo durante la resolución y evaluación del resultado. La retroalimentación formativa, en este contexto, actúa como información que calibra la percepción del estudiante sobre su propio desempeño, reduciendo la discrepancia entre la conciencia metacognitiva subjetiva y el desempeño real (Järvelä & Bannert, 2021).

El aula invertida, por su parte, crea una arquitectura de aprendizaje en la que las actividades de orden superior (aplicación, análisis, evaluación y creación) se concentran en el espacio presencial o sincrónico, mientras que la adquisición de contenidos básicos se desplaza al espacio autónomo. Esta inversión demanda un nivel significativo de autorregulación y metacognición en el espacio autónomo, donde el estudiante debe gestionar su tiempo, monitorear su comprensión y decidir cuándo necesita apoyo adicional (Molenaar et al., 2023). La IA puede jugar un papel particularmente valioso en este espacio, proporcionando retroalimentación adaptativa que suple la ausencia del docente.

La investigación reciente ha documentado que la integración de IA en metodologías activas puede crear lo que Järvelä et al. (2023) denominan bucles de regulación humano-IA (human-AI regulation loops), ciclos iterativos en los que las acciones del estudiante informan el comportamiento del sistema de IA, y la retroalimentación del sistema informa las acciones metacognitivas del estudiante. Estos bucles tienen el potencial de acelerar el desarrollo de la competencia metacognitiva, pero también de inhibirla si el diseño instruccional no los gestiona adecuadamente (Fan et al., 2025; Edwards et al., 2025).

### 2.5. Ética e inclusión en la implementación de IA educativa

El cuarto dominio teórico del modelo aborda la dimensión ética y de inclusión, frecuentemente relegada en la investigación sobre IA educativa pero crítica para una implementación responsable. Holmes & Tuomi (2022), en su revisión del estado del arte de la IA en educación, advierten que los sistemas actuales suelen incorporar concepciones del aprendizaje empobrecidas (frecuentemente reducidas al modelo conductista de adquisición de hechos) y que su despliegue masivo sin reflexión ética puede amplificar inequidades preexistentes en lugar de mitigarlas. Estudios subsecuentes con perspectivas de los actores educativos (Holmes et al., 2023; Cerratto et al., 2025) han identificado preocupaciones sustantivas en torno a la transparencia algorítmica, la protección de datos del aprendiz, la presencia de sesgos discriminatorios y la concentración del poder en plataformas comerciales.

Cualquier modelo teórico que proponga la integración sistemática de IA en procesos educativos debe, por tanto, incorporar estas dimensiones éticas no como apéndice sino como componente estructural. El modelo METACOGNITO adopta esta postura, incluyendo el diseño ético de la IA entre las variables moduladoras del ciclo y articulando un principio rector explícito: la IA debe operar como amplificador de la agencia del aprendiz, nunca como su sustituto.

## Metodología

### 3.1. Diseño y enfoque del estudio

Este trabajo adopta un enfoque cualitativo de naturaleza teórico-conceptual orientado a la construcción de un modelo teórico mediante la integración estructurada de evidencia bibliográfica reciente y de alto impacto. El diseño metodológico empleado es una revisión narrativa estructurada combinada con un procedimiento de síntesis basada en marcos (framework synthesis) (Schreiber & Cramer, 2024). Este tipo de revisión resulta apropiado cuando el objetivo del estudio no es agregar evidencia empírica para responder a una pregunta de eficacia (en cuyo caso correspondería una revisión sistemática con meta-análisis) sino mapear, clarificar y articular constructos teóricos para producir un modelo conceptual integrador.

La elección de este diseño se fundamenta en tres consideraciones. En primer lugar, las relaciones entre metacognición, retroalimentación formativa con IA y metodologías activas constituyen un campo conceptualmente fragmentado, en el que cada constructo cuenta con un cuerpo robusto de literatura

propio pero las articulaciones entre ellos son aún incipientes (Lan & Zhou, 2025; Tsakeni et al., 2025). En segundo lugar, la naturaleza emergente de la IA generativa en educación demanda síntesis teóricas que orienten la investigación empírica futura, antes de que esta se disperse en estudios desconectados. En tercer lugar, la construcción de modelos conceptuales sólidos requiere un proceso transparente y trazable de identificación, análisis e integración de la literatura existente (Peters et al., 2021).

### 3.2. Pregunta de investigación y marco conceptual

La pregunta de investigación que guió el proceso de revisión y síntesis se formuló mediante el marco PCC (Population, Concept, Context), recomendado para revisiones de naturaleza conceptual (Peters et al., 2021): ¿cómo puede la retroalimentación formativa mediada por inteligencia artificial (Concepto) andamiar el desarrollo de la conciencia metacognitiva (Concepto) de los estudiantes de educación superior (Población) en entornos de metodologías activas (Contexto)?

A partir de esta pregunta se establecieron cuatro dominios conceptuales que estructuraron tanto la búsqueda bibliográfica como el análisis posterior: (a) metacognición y aprendizaje autorregulado; (b) retroalimentación formativa mediada por inteligencia artificial, incluyendo la literatura sobre prompts metacognitivos; (c) engagement cognitivo y metodologías activas en educación superior; y (d) ética e inclusión en la implementación de IA educativa. La articulación entre estos cuatro dominios se planteó como el espacio teórico a explorar y modelar.

### 3.3. Estrategia de búsqueda y fuentes de información

La búsqueda bibliográfica se realizó en cuatro bases de datos de alto impacto: Web of Science (Core Collection), Scopus, ERIC (Education Resources Information Center) y PsycINFO. Estas bases fueron seleccionadas por su cobertura de revistas indexadas en los cuartiles superiores del Journal Citation Reports (JCR) y del SCImago Journal Rank (SJR) en las áreas de Educación, Psicología Educativa y Tecnología Educativa.

La estrategia de búsqueda combinó cuatro bloques temáticos mediante operadores booleanos. El primer bloque incluyó términos relativos a metacognición y autorregulación: "metacognition" OR "metacognitive awareness" OR "self-regulated learning" OR "metacognitive prompts". El segundo bloque agrupó términos asociados a la retroalimentación formativa mediada por IA: "formative feedback" OR "AI feedback" OR "automated feedback" OR "intelligent tutoring system" OR "large language model" OR "ChatGPT" OR "generative AI". El tercer bloque incluyó descriptores de metodologías activas y engagement: "active learning" OR "problem-based learning" OR "project-based learning" OR "flipped classroom" OR "cognitive engagement" OR "ICAP". El cuarto bloque incorporó términos de ética en IA educativa: "AI ethics" OR "responsible AI" OR "AI in education ethics". Los bloques se combinaron mediante el operador AND, permitiendo recuperar literatura en la intersección de al menos dos dominios.

El periodo de búsqueda priorizó publicaciones desde enero de 2021, con énfasis en literatura posterior a la masificación de los modelos generativos en educación a partir de 2022. Se incluyeron también obras fundacionales clásicas (Flavell, 1979; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Zimmerman, 2000; Chi & Wylie, 2014; Bannert, 2009; Fredricks et al., 2004; Reeve, 2013; Vygotsky, 1978) por su carácter de referencia obligatoria al describir los constructos centrales del modelo, siguiendo la práctica habitual en revisiones de síntesis conceptual (Schreiber & Cramer, 2024).

### 3.4. Criterios de inclusión y exclusión

Se establecieron criterios explícitos de inclusión y exclusión a fin de garantizar la calidad y pertinencia de la evidencia incorporada al análisis.

Criterios de inclusión: (a) artículos de investigación, revisiones sistemáticas, meta-análisis o artículos teórico-conceptuales publicados entre 2021 y la actualidad, con excepción de obras fundacionales esenciales para los constructos centrales; (b) publicaciones en revistas indexadas en Q1 según el JCR o el SJR en las categorías de Educación, Psicología Educativa o Tecnología Educativa; (c) trabajos escritos en inglés o español; (d) estudios que abordaran al menos dos de los cuatro dominios

conceptuales del marco PCC.

Criterios de exclusión: (a) literatura gris (tesis no publicadas, informes técnicos, documentos de trabajo); (b) capítulos de libros y actas de congresos no indexados en bases Q1; (c) editoriales, cartas al editor y comentarios; (d) estudios centrados exclusivamente en educación primaria o secundaria sin transferibilidad a educación superior; (e) trabajos cuyo enfoque tecnológico no involucrara IA propiamente dicha (por ejemplo, simples plataformas de gestión del aprendizaje sin componentes adaptativos o inteligentes).

### 3.5. Procedimiento de análisis y síntesis basada en marcos

Para el análisis del corpus se aplicó el procedimiento de *framework synthesis* recomendado por Schreiber & Cramer (2024) para revisiones de naturaleza conceptual. Este procedimiento consta de seis pasos: (1) construcción de un marco analítico inicial a partir de los constructos teóricos centrales identificados en el marco teórico; (2) familiarización con el corpus mediante lectura repetida de los artículos seleccionados; (3) codificación deductiva de los conceptos en función del marco inicial; (4) codificación inductiva de los conceptos emergentes que el marco inicial no anticipaba (entre los cuales destacó el constructo de pereza metacognitiva como tensión de diseño); (5) construcción de matrices conceptuales que organizan los hallazgos por dominio teórico y por relación entre dominios; y (6) síntesis interpretativa para la generación del modelo teórico final.

La codificación fue realizada por los autores siguiendo un proceso iterativo de discusión y consenso. Las divergencias interpretativas se resolvieron mediante revisión conjunta de las fuentes y reformulación de los códigos cuando fue necesario. Aunque este procedimiento no equivale a una medida formal de confiabilidad intercodificador del tipo kappa de Cohen, garantiza un nivel de triangulación apropiado para una síntesis conceptual.

### 3.6. Trazabilidad bibliográfica del modelo y consideraciones de rigor

El modelo teórico resultante se sometió a un proceso de trazabilidad bibliográfica explícita: cada componente estructural y cada relación entre componentes se documentó con respaldo bibliográfico verificable proveniente de al menos dos fuentes independientes del corpus. Importa subrayar que esta trazabilidad no constituye una validación empírica del modelo, sino una garantía de que las afirmaciones teóricas que lo conforman están ancladas en literatura revisada por pares. La validación empírica del modelo es una tarea explícitamente reservada para investigaciones futuras, como se discute en la sección de limitaciones.

El rigor metodológico del estudio se sustenta en cuatro criterios reconocidos en la investigación cualitativa: credibilidad, mediante la triangulación teórica entre marcos complementarios; transferibilidad, mediante la descripción detallada del proceso analítico que permite su replicación; dependabilidad, mediante el seguimiento sistemático del protocolo de *framework synthesis* (Schreiber & Cramer, 2024); y confirmabilidad, mediante la trazabilidad explícita de cada componente del modelo a referencias bibliográficas verificables. El estudio no requirió aprobación ética dado que no involucra recolección de datos primarios con seres humanos.

## Resultados: el modelo METACOGNITO

La aplicación del procedimiento de síntesis basada en marcos al corpus de estudios incluidos permitió identificar los componentes conceptuales, las relaciones entre dominios y la estructura procesual que articulan el modelo teórico aquí propuesto. Los resultados se presentan en seis apartados: la fundamentación del modelo, sus componentes estructurales, su ciclo procesual, su comparación con marcos clásicos de SRL, las proposiciones contrastables que se derivan del modelo y, finalmente, el mapa conceptual que integra los hallazgos.

### 4.1. Presentación y fundamentación del modelo

Como producto principal de la síntesis, se propone el modelo METACOGNITO. El nombre del modelo es una denominación coined que evoca la articulación entre metacognición y cognición situada en entornos mediados por IA, y no opera como un acrónimo en sentido estricto. El modelo es de

naturaleza procesual y sistémica: procesual porque describe un ciclo de fases secuenciales e iterativas, y sistémica porque postula relaciones de retroalimentación y dependencia mutua entre sus componentes.

El modelo METACOGNITO se fundamenta en cuatro supuestos teóricos centrales que emergieron de manera consistente del análisis del corpus:

La conciencia metacognitiva no es un estado fijo sino una capacidad dinámica que se desarrolla a través de la práctica reflexiva sostenida en contextos auténticos de aprendizaje (Tsakeni et al., 2025; Wang et al., 2022).

La retroalimentación formativa mediada por IA puede actuar como andamiaje metacognitivo externo cuando su diseño está orientado explícitamente a activar los procesos de monitoreo y regulación interna del aprendiz, y no únicamente a corregir errores (Bellhäuser et al., 2023; Banihashem et al., 2025; Bannert, 2009).

Las metodologías activas constituyen el contexto pedagógico óptimo para que la articulación IA-metacognición produzca efectos de aprendizaje profundo y transferible, en la medida en que activan modos de engagement cognitivo constructivo o interactivo (Chi & Wylie, 2014; Chiu et al., 2023).

El uso ético, transparente e inclusivo de la IA es una condición de posibilidad, y no un requisito accesorio, para que el modelo produzca efectos positivos sobre el aprendizaje (Holmes & Tuomi, 2022; Cerratto et al., 2025).

#### 4.2. Componentes estructurales del modelo

El modelo METACOGNITO articula tres planos interrelacionados:

Plano A. El aprendiz metacognitivo: describe las dimensiones internas del estudiante que el modelo busca desarrollar.

Plano B. El sistema de retroalimentación con IA: describe los mecanismos tecnológicos y pedagógicos de la retroalimentación formativa mediada por IA, incluyendo el conector metacognitivo como elemento original del modelo.

Plano C. El entorno pedagógico activo: describe las características del contexto instruccional que habilitan la articulación de los planos A y B.

##### 4.2.1. Plano A. El aprendiz metacognitivo

Este plano comprende tres dimensiones metacognitivas interrelacionadas, basadas en la síntesis de Azevedo et al. (2022), Järvelä & Bannert (2021) y Lan & Zhou (2025):

Conocimiento metacognitivo: el saber declarativo, procedimental y condicional que el aprendiz posee sobre su propio funcionamiento cognitivo (qué sabe, cómo aprende mejor, cuándo aplicar qué estrategia).

Monitoreo metacognitivo: la capacidad de seguimiento en tiempo real del propio estado de comprensión y progreso hacia los objetivos de aprendizaje.

Control metacognitivo: la capacidad de ajustar activamente las estrategias de aprendizaje en función de los resultados del monitoreo.

##### 4.2.2. Plano B. El sistema de retroalimentación con IA y el conector metacognitivo

Este plano engloba los componentes funcionales del sistema de IA que median la retroalimentación formativa, organizados en tres capas:

Capa diagnóstica: el sistema de IA analiza continuamente las producciones del estudiante, sus patrones de interacción y sus respuestas a retroalimentaciones previas, construyendo un modelo dinámico del aprendiz (Molenaar et al., 2023; Lim et al., 2023).

Capa generativa: el sistema produce retroalimentación personalizada en los cuatro niveles del

modelo de Hattie & Timperley (2007), con énfasis en los niveles de proceso y autorregulación.

Capa metacognitiva: el sistema incorpora explícitamente conectores metacognitivos diseñados para activar el monitoreo y el control metacognitivo del estudiante, más allá de la corrección del error.

El conector metacognitivo es la contribución original más significativa del modelo. Se define como un elemento de diseño de la retroalimentación que acopla cada pieza de información formativa entregada al estudiante con una pregunta o indicación reflexiva diseñada para ser procesada a través de su filtro metacognitivo. El conector hereda la lógica de los prompts metacognitivos clásicos (Bannert, 2009; Guo, 2022; Wong et al., 2021) pero se diferencia de ellos en tres dimensiones específicas:

Acoplamiento estructural al ciclo de retroalimentación: a diferencia de los prompts metacognitivos clásicos, que típicamente operan como intervenciones aisladas en momentos predeterminados, el conector se adhiere estructuralmente a cada pieza de retroalimentación formativa, formando una unidad indisociable de información-más-reflexión.

Intencionalidad multinivel: el conector está diseñado para articular simultáneamente los cuatro niveles de retroalimentación de Hattie & Timperley (2007) con las tres fases del SRL de Zimmerman (2000), garantizando que la reflexión metacognitiva activada incluya tanto la dimensión de la tarea como la del proceso, la autorregulación y, eventualmente, el yo.

Adaptatividad dinámica: el conector se ajusta al perfil metacognitivo cambiante del estudiante a partir del modelo del aprendiz mantenido por la capa diagnóstica del sistema, lo cual no es factible en los sistemas de prompts predefinidos basados en reglas.

#### 4.2.3. Plano C. El entorno pedagógico activo

Este plano describe las características estructurales del entorno instruccional que habilitan la integración efectiva de los planos A y B:

Autenticidad de la tarea: las tareas deben presentar complejidad cognitiva real que demande planificación, monitoreo y evaluación por parte del estudiante.

Agencia estudiantil: el diseño instruccional debe ofrecer espacios de decisión y autogestión que requieran del ejercicio activo de la regulación metacognitiva, conectando con el constructo de engagement agéntico (Reeve, 2013).

Modo de engagement constructivo o interactivo: siguiendo el marco ICAP (Chi & Wylie, 2014), el entorno debe favorecer modos de engagement en los que el estudiante genere nuevas inferencias o coconstruya conocimiento con otros agentes (humanos o IA).

Dialogicidad: el entorno debe posibilitar el diálogo reflexivo, tanto entre pares como con el sistema de IA, como medio de externalización y calibración metacognitiva.

Temporalidad cíclica: las actividades deben organizarse en ciclos iterativos que permitan al estudiante experimentar el efecto de sus ajustes metacognitivos sobre sus resultados de aprendizaje.

#### 4.3. El ciclo procesual del modelo METACOGNITO

El corazón dinámico del modelo es un ciclo de cinco fases que describe el proceso mediante el cual la retroalimentación formativa con IA activa y desarrolla la conciencia metacognitiva del estudiante en un entorno de metodología activa.

##### Fase 1. Activación

El estudiante se enfrenta a una tarea auténtica dentro de la metodología activa (un problema, un proyecto, un caso). Antes de comenzar, el sistema de IA proporciona un andamiaje de planificación metacognitiva mediante preguntas guía del tipo "¿qué ya sabes sobre este tema?", "¿qué estrategias usarás?", "¿cómo sabrás que has tenido éxito?", que activa el conocimiento metacognitivo previo del estudiante y establece un punto de referencia para el monitoreo posterior. Esta fase corresponde a la etapa de previsión (forethought) del modelo de autorregulación de Zimmerman (2000) e incorpora explícitamente el monitoreo anticipatorio descrito por Molenaar et al. (2023).

##### Fase 2. Retroalimentación dinámica

Durante la ejecución de la tarea, el sistema de IA genera retroalimentación formativa en tiempo real y diferida. La retroalimentación está estratificada en los cuatro niveles de Hattie & Timperley (2007), pero la novedad del modelo reside en que cada tipo de retroalimentación va acompañado de un conector metacognitivo que solicita al estudiante procesar la información a través de su filtro metacognitivo. Esta fase operacionaliza los principios de Nicol & Macfarlane-Dick (2006) en el contexto específico de la IA generativa, a la luz de los hallazgos de Urban et al. (2025) y Xu et al. (2025) sobre el papel de la alfabetización metacognitiva en el procesamiento de la retroalimentación de IA.

### Fase 3. Reflexión metacognitiva

Tras recibir la retroalimentación, el estudiante participa en una actividad estructurada de reflexión metacognitiva, facilitada pero no determinada por el sistema de IA. Esta reflexión puede tomar la forma de un diario de aprendizaje, una autoevaluación comparativa, o una discusión en pares guiada por preguntas metacognitivas generadas por la IA. El objetivo de esta fase es transformar la retroalimentación externa en regulación interna: que el estudiante no solo corrija el error, sino que actualice su modelo mental de sí mismo como aprendiz (Wang et al., 2022). Esta fase es la más vulnerable al riesgo de pereza metacognitiva descrito por Fan et al. (2025) y Zhang et al. (2024), razón por la cual su diseño debe incluir andamiajes que exijan esfuerzo cognitivo genuino del estudiante.

### Fase 4. Ajuste regulatorio

Con base en su reflexión metacognitiva, el estudiante realiza ajustes explícitos en sus estrategias de aprendizaje, sus metas o su gestión del tiempo. El sistema de IA monitorea estos ajustes (la calidad de la planificación revisada, la coherencia entre el ajuste y el problema identificado, el tiempo invertido en la reflexión) y calibra su modelo del aprendiz en consecuencia. Cuando el sistema detecta patrones de ajuste inefectivos o estrategias persistentemente disfuncionales, genera retroalimentación de segundo orden, esto es, retroalimentación sobre el proceso de ajuste metacognitivo mismo, que opera en el nivel más profundo de autorregulación descrito por Hattie & Timperley (2007). Esta fase corresponde al bucle de regulación humano-IA (Järvelä et al., 2023), donde la IA actúa no como fuente de respuestas sino como espejo regulatorio del proceso metacognitivo del estudiante.

### Fase 5. Transferencia y consolidación

La fase final del ciclo busca garantizar que los aprendizajes metacognitivos adquiridos durante las fases anteriores se transfieran a nuevos contextos y tareas. El sistema de IA facilita esta transferencia mediante la generación de conexiones explícitas entre las estrategias que resultaron efectivas en la tarea actual y la arquitectura general de competencias metacognitivas del estudiante. La consolidación incluye la generación de un perfil metacognitivo dinámico por parte del sistema de IA, que el estudiante puede revisar y utilizar como herramienta de planificación en contextos futuros (Lan & Zhou, 2025; Bellhäuser et al., 2023). Esta fase cierra el ciclo y establece el punto de partida de una nueva iteración.

#### 4.3.1. Ejemplos de conectores metacognitivos por nivel de retroalimentación

La Tabla 1 ilustra el constructo del conector metacognitivo mediante ejemplos para cada uno de los cuatro niveles de retroalimentación de Hattie & Timperley (2007), que pueden ser adaptados por el sistema de IA al perfil del aprendiz y al tipo de tarea.

| Nivel de retroalimentación | Retroalimentación de la IA (ejemplo)                                  | Conector metacognitivo asociado (ejemplo)                                                                                       |
|----------------------------|-----------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Tarea                      | "La fórmula aplicada en el paso 3 no corresponde al tipo de problema" | "¿Qué señal del enunciado del problema te llevó a elegir esa fórmula? ¿Qué pista habrías necesitado para detectar el desajuste" |

| Nivel de retroalimentación | Retroalimentación de la IA (ejemplo)                                                                 | Conector metacognitivo asociado (ejemplo)                                                                                               |
|----------------------------|------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
|                            | planteado."                                                                                          | antes?"                                                                                                                                 |
| Proceso                    | "Tu argumento avanza directamente a la conclusión sin examinar contraargumentos."                    | "¿Tu plan inicial contemplaba abordar contraargumentos? ¿En qué momento del proceso decidiste omitirlos y qué te llevó a esa decisión?" |
| Autorregulación            | "Llevas 40 minutos en este apartado y has cambiado de estrategia tres veces sin completar ninguna."  | "¿Qué te dice este patrón sobre cómo gestionas la incertidumbre? ¿Qué harías distinto si volvieras a empezar la sesión?"                |
| Yo (self)                  | (Nivel evitado por defecto en el modelo, salvo refuerzos explícitos del esfuerzo y la persistencia.) | "¿Qué cosas has hecho hoy que muestran tu compromiso con esta tarea, independientemente del resultado obtenido?"                        |

Tabla 1. Ejemplos de conectores metacognitivos por nivel de retroalimentación.

#### 4.4. Comparación con modelos clásicos de aprendizaje autorregulado

Una de las contribuciones del modelo METACOGNITO es ofrecer una correspondencia explícita entre sus cinco fases y las fases descritas en los modelos clásicos de SRL. La Tabla 2 presenta esta correspondencia con los modelos de Zimmerman (2000), Pintrich y el modelo COPES de Winne, según se sintetiza en revisiones contemporáneas (Molenaar et al., 2023; Fan et al., 2023).

Tabla 2. Correspondencia entre las fases del modelo METACOGNITO y los modelos clásicos de aprendizaje autorregulado.

| Fase METACOGNITO                      | Zimmerman (2000)                                 | Pintrich (2000)                      | Winne / COPES                                |
|---------------------------------------|--------------------------------------------------|--------------------------------------|----------------------------------------------|
| Fase 1. Activación                    | Previsión (forethought)                          | Planificación y activación           | Definición de la tarea                       |
| Fase 2. Retroalimentación dinámica    | Ejecución (performance)                          | Monitoreo                            | Aplicación de tácticas y estrategias         |
| Fase 3. Reflexión metacognitiva       | Ejecución / Autorreflexión                       | Control                              | Adaptación y metacognición                   |
| Fase 4. Ajuste regulatorio            | Autorreflexión (self-reflection)                 | Reacción y reflexión                 | Adaptación y metacognición                   |
| Fase 5. Transferencia y consolidación | Autorreflexión que retroalimenta nueva previsión | Reflexión que reactiva planificación | Cambio en estándares y conocimiento de tarea |

#### 4.5. Variables moduladoras y proposiciones contrastables

El análisis del corpus permitió identificar un conjunto de variables contextuales e individuales

Retroalimentación formativa mediada por inteligencia artificial y conciencia metacognitiva: un modelo teórico para las pedagogías activas AI-Driven Formative Feedback and Metacognitive Awareness: A Theoretical Model for Active Pedagogies.

que modulan la efectividad del ciclo procesual. A diferencia de descripciones meramente narrativas, estas variables se formulan a continuación como proposiciones contrastables, con el fin de orientar la investigación empírica futura sobre el modelo:

**Proposición 1 (alfabetización en retroalimentación).** La alfabetización en retroalimentación del estudiante moderará positivamente la relación entre la calidad de la retroalimentación generada por IA en la Fase 2 y la profundidad de la reflexión metacognitiva activada en la Fase 3, controlando por el nivel de conocimiento previo (Carless & Winstone, 2023; Lipnevich & Smith, 2022).

**Proposición 2 (alfabetización en IA).** La alfabetización en IA del estudiante moderará la confiabilidad percibida de la retroalimentación generada artificialmente, de modo que estudiantes con mayor alfabetización en IA aceptarán o rechazarán críticamente la retroalimentación con base en su pertinencia, mientras que estudiantes con baja alfabetización en IA tenderán al sobreajuste o al rechazo acrítico (Urban et al., 2025; Chiu et al., 2023).

**Proposición 3 (orientación motivacional).** Los estudiantes con orientación hacia el dominio (mastery orientation) utilizarán los conectores metacognitivos de manera más profunda y producirán mayores ganancias en conciencia metacognitiva que los estudiantes con orientación hacia el rendimiento (performance orientation), independientemente de la calidad técnica del sistema de IA (Fan et al., 2023).

**Proposición 4 (modo de engagement).** La efectividad del ciclo METACOGNITO sobre el desarrollo de la conciencia metacognitiva será mayor en entornos pedagógicos que activen modos de engagement constructivos o interactivos (en términos del marco ICAP de Chi & Wylie, 2014) que en entornos que activen modos pasivos o meramente activos, manteniendo constantes el resto de variables del modelo.

**Proposición 5 (diseño ético).** La transparencia algorítmica, la protección de datos y la ausencia de sesgos discriminatorios en el sistema de IA constituyen condiciones de posibilidad del modelo: en su ausencia, los efectos esperados sobre el aprendizaje no se materializarán o lo harán de manera desigual entre subgrupos de estudiantes (Holmes & Tuomi, 2022; Cerratto et al., 2025).

#### 4.6. Mapa conceptual del modelo

La Figura 1 sintetiza visualmente los hallazgos. Su valor analítico, más allá de la descripción de su estructura, reside en tres insights que el mapa hace evidentes. En primer lugar, las cinco fases del ciclo no son simétricas: la Fase 3 (reflexión metacognitiva) concentra la mayor vulnerabilidad ante el riesgo de pereza metacognitiva, lo que la convierte en el punto crítico de diseño instruccional. En segundo lugar, las variables moduladoras no operan sobre fases específicas sino transversalmente sobre todo el ciclo, lo que implica que cualquier intervención educativa basada en el modelo debe abordarlas como condiciones de borde antes que como elementos secundarios. En tercer lugar, el flujo F5 → F1 (línea punteada) representa la dimensión iterativa que distingue al modelo de los esquemas lineales de retroalimentación: el aprendizaje metacognitivo no termina en una tarea sino que se acumula a través de ciclos sucesivos.

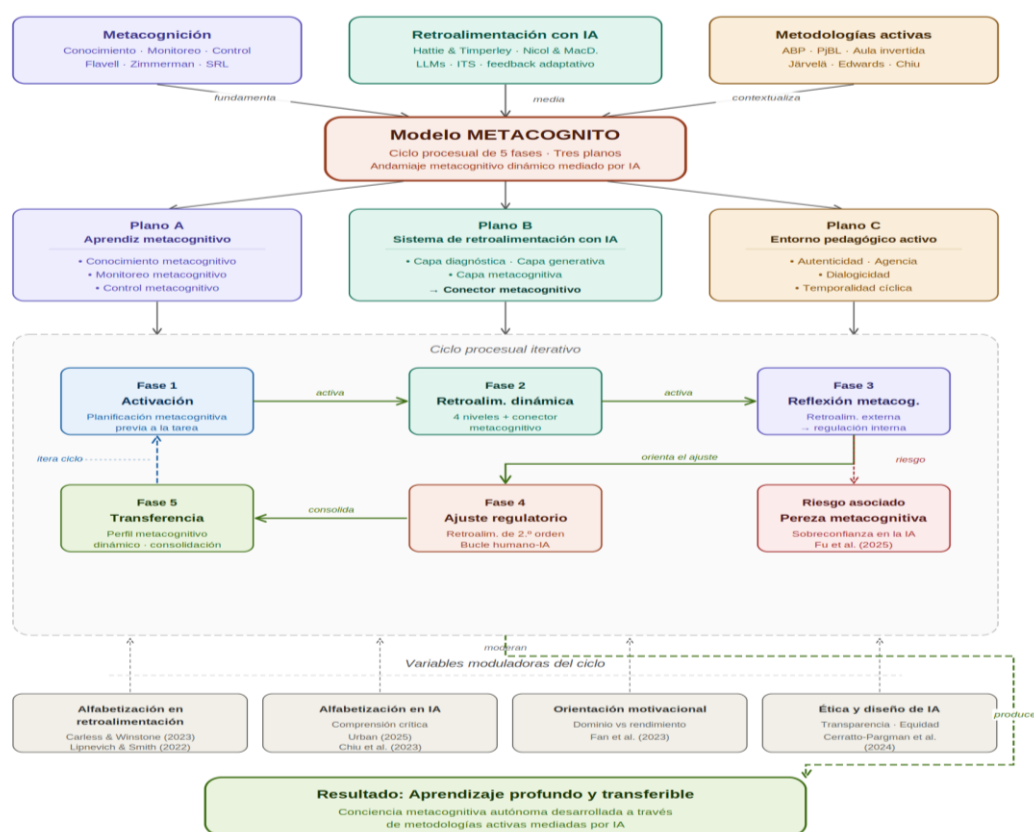


Figura 1. Mapa conceptual del modelo METACOGNITO. Elaboración propia.

## Discusión

### 5.1. Aportes teóricos al campo

El modelo METACOGNITO aporta al campo educativo un marco integrador que trasciende los enfoques parciales predominantes en la literatura. Mientras la mayoría de los estudios existentes abordan la IA y la metacognición como variables separadas o examinan su relación de manera descriptiva (Tsakeni et al., 2025; Lan & Zhou, 2025), el modelo METACOGNITO las sitúa en una relación de articulación procesual explícita dentro de un contexto pedagógico específico, anclado en el marco ICAP (Chi & Wylie, 2014).

La originalidad más significativa del modelo reside en la incorporación del conector metacognitivo como elemento de diseño de la retroalimentación. La literatura ha establecido con solidez que la retroalimentación eficaz opera en niveles de proceso y autorregulación (Hattie & Timperley, 2007), y que la retroalimentación de la IA puede llegar a ser comparable en calidad a la humana cuando está adecuadamente diseñada (Bellhäuser et al., 2023; Banihashem et al., 2025). Igualmente, la literatura sobre prompts metacognitivos ha demostrado tamaños de efecto positivos sobre la SRL y los resultados de aprendizaje (Bannert, 2009; Guo, 2022; Wong et al., 2021). No obstante, la mayoría de los sistemas actuales de retroalimentación con IA no incorporan explícitamente elementos diseñados para activar la reflexión metacognitiva del estudiante; la corrección del error y la orientación hacia la mejora son sus objetivos principales, pero no la activación de los procesos internos de monitoreo y regulación cognitiva. El conector metacognitivo intenta subsanar esta brecha integrando la lógica de los prompts metacognitivos con la arquitectura de la retroalimentación generativa adaptativa.

### 5.2. La paradoja de la pereza metacognitiva y su tratamiento en el modelo

El modelo incorpora la advertencia teórica sobre la pereza metacognitiva identificada por Fan et al. (2025) y Zhang et al. (2024) como una tensión de diseño fundamental que cualquier sistema de retroalimentación con IA debe resolver. La paradoja es real: la facilidad y la inmediatez de la retroalimentación generada por IA pueden crear condiciones de dependencia que erosionan

precisamente la capacidad metacognitiva que se pretende desarrollar.

El modelo METACOGNITO propone resolver esta paradoja a través del principio de andamiaje con desfase progresivo: el sistema de IA comienza ofreciendo andamiaje metacognitivo explícito y elaborado, pero va retirándolo gradualmente a medida que el estudiante demuestra mayor autonomía metacognitiva, siguiendo el principio de la zona de desarrollo próximo (ZDP) de Vygotsky (1978). Este principio se operacionaliza, en el plano técnico, mediante el modelo del aprendiz mantenido por la capa diagnóstica del sistema, que ajusta dinámicamente el nivel de elaboración de los conectores metacognitivos en función del desempeño metacognitivo demostrado por el estudiante.

### 5.3. Consideraciones éticas, de inclusión y culturales

Desde una perspectiva crítica, el modelo reconoce que la implementación de sistemas de IA en contextos educativos no es un proceso técnicamente neutral: entraña decisiones de diseño con implicaciones éticas, políticas y pedagógicas profundas (Holmes & Tuomi, 2022; Holmes et al., 2023). La dependencia de plataformas comerciales de IA, los sesgos potenciales en los algoritmos de retroalimentación, la protección de los datos de aprendizaje de los estudiantes y el riesgo de amplificar desigualdades preexistentes son dimensiones que cualquier implementación del modelo debe abordar explícitamente (Cerratto et al., 2025; Chan, 2023). El modelo METACOGNITO adopta como principio rector que la IA debe funcionar en todo momento como un amplificador de la agencia del aprendiz, nunca como su sustituto.

En relación con la inclusión, el modelo reconoce que las dimensiones del aprendiz metacognitivo (Plano A) pueden manifestarse de manera distinta en estudiantes con discapacidades de aprendizaje, condiciones neurodivergentes (TDAH, autismo, dislexia) o necesidades educativas específicas. La adaptación del modelo a estos perfiles requiere ajustes específicos: por ejemplo, el conector metacognitivo puede beneficiarse de formatos multimodales para estudiantes con dislexia, de tiempos extendidos para estudiantes con TDAH o de estructuras explícitas y predecibles para estudiantes en el espectro autista. Estos ajustes no son contrarios al modelo, sino expresiones de su principio de adaptatividad dinámica.

Desde una perspectiva cultural, el modelo se ha desarrollado predominantemente con referencia a literatura producida en contextos académicos occidentales individualistas. En contextos culturales diferentes, particularmente aquellos con tradiciones colectivistas o de herencia confuciana, las dinámicas metacognitivas y de retroalimentación pueden presentar configuraciones distintas, especialmente en lo que se refiere al rol de la autoridad docente, las normas de comunicación crítica entre pares y las concepciones de éxito académico. La aplicación del modelo a estos contextos requiere validaciones específicas que se proponen como línea futura de investigación.

### 5.4. Implicaciones para el diseño instruccional, la formación docente y la política educativa

En términos de diseño instruccional, el modelo sugiere que la integración de IA en metodologías activas debe ser guiada por principios de diseño metacognitivo explícito, no simplemente por la disponibilidad tecnológica. Los desarrolladores de sistemas educativos deberían incorporar conectores metacognitivos como elemento estándar de los flujos de retroalimentación, en lugar de tratarlos como funcionalidades opcionales.

En términos de formación docente, el modelo implica que los profesores necesitan desarrollar competencias específicas de alfabetización en IA y en retroalimentación para poder diseñar y gestionar los conectores metacognitivos de manera efectiva. La formación docente debería incluir explícitamente la dimensión metacognitiva del diseño de tareas y la articulación entre tarea, retroalimentación y reflexión.

En términos de política educativa, el modelo subraya la necesidad de marcos regulatorios que garanticen el uso ético de los datos de aprendizaje, la equidad en el acceso a tecnologías de retroalimentación con IA y la transparencia de los algoritmos utilizados en sistemas educativos. La inversión en infraestructura tecnológica debe acompañarse necesariamente de inversión en marcos

éticos institucionales.

## Conclusiones, limitaciones y futuras líneas de investigación

### 6.1. Conclusiones

Este artículo ha desarrollado el modelo METACOGNITO, un marco teórico-conceptual que integra de manera sistemática cuatro dominios de la investigación educativa contemporánea: la metacognición y el aprendizaje autorregulado, la retroalimentación formativa mediada por inteligencia artificial (incluyendo la literatura sobre prompts metacognitivos), las metodologías activas centradas en el estudiante anclado en el marco ICAP, y la ética de la IA en educación. El modelo describe un ciclo procesual de cinco fases (Activación, Retroalimentación Dinámica, Reflexión Metacognitiva, Ajuste Regulatorio y Transferencia) en el que los sistemas de IA actúan como andamiajes dinámicos que amplían el potencial metacognitivo del aprendiz en entornos pedagógicos activos.

Las principales contribuciones del modelo son: (1) la conceptualización del conector metacognitivo como elemento de diseño que se diferencia de los prompts metacognitivos clásicos por su acoplamiento al ciclo de retroalimentación, su intencionalidad multinivel y su adaptatividad; (2) la articulación explícita entre los ciclos de retroalimentación formativa, las fases del aprendizaje autorregulado y los modos de engagement cognitivo del marco ICAP; (3) la formulación de cinco proposiciones contrastables que orientan la investigación empírica futura; (4) la incorporación de la pereza metacognitiva como tensión de diseño y del andamiaje con desfase progresivo como principio resolutorio; y (5) la integración estructural de la dimensión ética y de inclusión como condición de posibilidad del modelo.

### 6.2. Limitaciones del estudio

El estudio presenta limitaciones que conviene reconocer explícitamente. La primera y más significativa es la naturaleza exclusivamente teórica del modelo: las relaciones propuestas entre sus componentes son plausibles a la luz de la literatura existente, pero requieren validación empírica sistemática. La trazabilidad bibliográfica establecida en el procedimiento metodológico garantiza el anclaje del modelo en evidencia revisada por pares, pero no equivale a una prueba de su efectividad o de la corrección de sus relaciones causales propuestas.

Una segunda limitación es el sesgo de publicación introducido por la restricción del corpus a publicaciones indexadas en Q1. Esta restricción privilegia la calidad pero excluye literatura potencialmente relevante publicada en revistas Q2 a Q4, en literatura gris, en idiomas distintos al inglés y al español, y en contextos del Sur Global donde la producción académica indexada en Q1 sigue siendo proporcionalmente menor. Una tercera limitación, conectada con la anterior, es el sesgo cultural ya discutido en la sección 5.3.

Una cuarta limitación es metodológica: a diferencia de una revisión sistemática con protocolo PRISMA registrado prospectivamente, la revisión narrativa estructurada empleada en este trabajo presenta un grado mayor de discrecionalidad en la selección y el análisis de fuentes. Aunque el procedimiento de framework synthesis aplicado mitiga este riesgo mediante la trazabilidad explícita de cada componente del modelo, no lo elimina por completo.

Una quinta limitación reside en la operacionalización tecnológica del modelo. El conector metacognitivo, tal como se describe, requiere capacidades técnicas (modelo dinámico del aprendiz, generación adaptativa de preguntas reflexivas, monitoreo en tiempo real del ajuste metacognitivo) que solo están parcialmente disponibles en los sistemas comerciales de IA actuales. La implementación práctica del modelo dependerá de avances tecnológicos que escapen al alcance de este trabajo.

### 6.3. Futuras líneas de investigación

Las futuras líneas de investigación que se derivan del modelo incluyen, en primer lugar, estudios experimentales que evalúen la efectividad del modelo en contextos reales de metodologías activas, con diseños comparativos entre intervenciones con conectores metacognitivos y condiciones de retroalimentación de IA estándar. En segundo lugar, investigación de diseño (design-based research)

que permita iterar y refinar el modelo en respuesta a evidencia de implementación. En tercer lugar, estudios de análisis de aprendizaje (learning analytics) que permitan operacionalizar y medir el ciclo METACOGNITO con datos reales de trazas de interacción.

En cuarto lugar, se proponen estudios de validación cultural del modelo en contextos no occidentales y con poblaciones culturalmente diversas. En quinto lugar, investigaciones específicas sobre la adaptación del modelo a estudiantes con discapacidades de aprendizaje y condiciones neurodivergentes. En sexto lugar, estudios cualitativos en profundidad sobre la experiencia subjetiva del estudiante al interactuar con conectores metacognitivos, con énfasis en los mecanismos por los cuales la pereza metacognitiva emerge o se evita. En séptimo lugar, investigación sobre las condiciones de formación docente necesarias para implementar el modelo de manera efectiva en contextos institucionales reales..

## References

- Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., Cloude, E., Dever, D., Wiedbusch, M., Wortha, F., & Cerezo, R. (2022). Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology*, 13, 813632. <https://doi.org/10.3389/fpsyg.2022.813632>
- Banihashem, S. K., Gašević, D., Noroozi, O., Jarodzka, H., Joosten-ten Brinke, D., & Drachsler, H. (2025). Optimizing formative assessment with learning analytics. *Review of Educational Research*. Publicación anticipada en línea. <https://doi.org/10.3102/00346543251370753>
- Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie*, 23(2), 139-145. <https://doi.org/10.1024/1010-0652.23.2.139>
- Bellhäuser, H., Dignath, C., & Theobald, M. (2023). Daily automated feedback enhances self-regulated learning: A longitudinal randomized field experiment. *Frontiers in Psychology*, 14, 1125873. <https://doi.org/10.3389/fpsyg.2023.1125873>
- Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1), 150-163. <https://doi.org/10.1080/13562517.2020.1782372>
- Cerratto, T., McGrath, C., & Milrad, M. (2025). Towards responsible AI in education: Challenges and implications for research and practice. *Computers and Education: Artificial Intelligence*, 7, 100345. <https://doi.org/10.1016/j.caeai.2024.100345>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(38). <https://doi.org/10.1186/s41239-023-00408-3>
- Chi, M., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243. <https://doi.org/10.1080/00461520.2014.965823>
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., Roberts, A.-S., & Järvelä, S. (2025). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology*, 56, 712-733. <https://doi.org/10.1111/bjet.13534>
- Fan, Y., Rakovic, M., van derGraaf, J., Lim, L., Singh, S., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2023). Towards a fuller picture: Triangulation and integration of the measurement of self-regulated learning based on trace and think aloud data. *Journal of Computer Assisted Learning*, 39(4), 1303-1324. <https://doi.org/10.1111/jcal.12801>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(1), 489-530. <https://doi.org/10.1111/bjet.13544>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry.

American Psychologist, 34(10), 906-911. <https://doi.org/10.1037/0003-066X.34.10.906>

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109. <https://doi.org/10.3102/00346543074001059>

Guo, L. (2022). Using metacognitive prompts to enhance self-regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments. *Journal of Computer Assisted Learning*, 38(3), 811-832. <https://doi.org/10.1111/jcal.12650>

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570. <https://doi.org/10.1111/ejed.12533>

Holmes, W., Iniesto, F., Anastopoulou, S., & Boticario, J. G. (2023). Stakeholder perspectives on the ethics of AI in distance-based higher education. *International Review of Research in Open and Distributed Learning*, 24(2), 96-117. <https://doi.org/10.19173/irrodl.v24i2.6089>

Järvelä, S., & Bannert, M. (2021). Temporal and adaptive processes of regulated learning: What can multimodal data tell? *Learning and Instruction*, 72, 101268. <https://doi.org/10.1016/j.learninstruc.2019.101268>

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, 54(5), 1057-1076. <https://doi.org/10.1111/bjet.13325>

Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Science of Learning*, 10, 21. <https://doi.org/10.1038/s41539-025-00319-0>

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, 139, 107547. <https://doi.org/10.1016/j.chb.2022.107547>

Lipnevich, A. A., & Smith, J. K. (2022). Student-feedback interaction model: Revised. *Studies in Educational Evaluation*, 75, 101208. <https://doi.org/10.1016/j.stueduc.2022.101208>

Molenaar, I., de Mooij, S., Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, 139, 107540. <https://doi.org/10.1016/j.chb.2022.107540>

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>

Peters, M. D. J., Marnie, C., Colquhoun, H., Garritty, C. M., Hempel, S., Horsley, T., Langlois, E. V., Lillie, E., O'Brien, K. K., Tunçalp, Ö., Wilson, M. G., Zarin, W., & Tricco, A. C. (2021). Scoping reviews: Reinforcing and advancing the methodology and application. *Systematic Reviews*, 10(1), 263. <https://doi.org/10.1186/s13643-021-01821-3>

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, 105(3), 579-595. <https://doi.org/10.1037/a0032690>

Schreiber, F., & Cramer, C. (2024). Towards a conceptual systematic review: Proposing a methodological framework. *Educational Review*, 75(7), 1458-1479. <https://doi.org/10.1080/00131911.2022.2116561>

Tsakeni, M., Nwafor, S. C., Mosia, M., & Egara, F. O. (2025). Mapping the scaffolding of metacognition and learning by AI tools in STEM classrooms: A bibliometric-systematic review approach (2005-2025). *Journal of Intelligence*, 13(11), 148. <https://doi.org/10.3390/jintelligence13110148>

Urban, M., Brom, C., Lukavský, J., Děchtěrenko, F., Hein, V., Svacha, F., Kmoníčková, P., & Urban, K. (2025). "ChatGPT can make mistakes. Check important info." Epistemic beliefs and metacognitive accuracy in students' integration of ChatGPT content into academic writing. *British Journal of Educational Technology*, 56, 1897-1918. <https://doi.org/10.1111/bjet.13591>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, F., Meng, F., Liu, S., Wang, S., Pan, L., & Lin, Z. (2022). Understanding learners' metacognition of online teacher feedback amid COVID-19: A case study in a university livestream instruction context. *Frontiers in*

Retroalimentación formativa mediada por  
inteligencia artificial y conciencia metacognitiva: un  
modelo teórico para las pedagogías activas AI-  
Driven Formative Feedback and Metacognitive  
Awareness: A Theoretical Model for Active  
Pedagogies.

Psychology, 13, 861845. <https://doi.org/10.3389/fpsyg.2022.861845>

Wong, J., Baars, M., de Koning, B. B., & Paas, F. (2021). Examining the use of prompts to facilitate self-regulated learning in massive open online courses. *Computers in Human Behavior*, 115, 106596. <https://doi.org/10.1016/j.chb.2020.106596>

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, 56, 1842–1863. <https://doi.org/10.1111/bjet.13599>

Zhang, S., Zhao, X., Zhou, T., & Kim, J. H. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. *International Journal of Educational Technology in Higher Education*, 21(34). <https://doi.org/10.1186/s41239-024-00467-0>

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. En M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>.