

Inteligencia artificial en educación superior: arquitecturas agénticas, feedback y enfoques pedagógicos emergentes. Una revisión de alcance 2024-2026

Artificial Intelligence in Higher Education: Agentic Architectures, Feedback, and Emerging Pedagogical Approaches. A Scoping Review, 2024–2026

<https://doi.org/10.47606/ACVEN/PH0588>

José Eugenio Chafloque Capuñay^{1*}

<https://orcid.org/0009-0002-4642-1540>

jose.chafloque@unprg.edu.pe

Mariluz Herrera Cervantes²

<https://orcid.org/0009-0002-3360-6247>

herrera.mariluz@usfx.bo

Guadalupe Badillo Márquez³

<https://orcid.org/0000-0002-0364-9329>

badillo.marquez.g@bine.mx

Elsa Dolores Hernández García⁴

<https://orcid.org/0009-0002-5839-2118>

elsa.hernandez@adenuniversity.edu.pa

Germán Vicente Garay-Flores⁵

<https://orcid.org/0000-0002-7118-6477>

ggarayfl01@ucvvirtual.edu.pe

Gladys Huarachi Chuquimia⁶

<https://orcid.org/0000-0003-3802-2574>

ghuarachich@unjbq.edu.pe

Recibido: 03/07/2026

Aceptado: 22/09/2026

RESUMEN

Recientes trabajos sobre IA en la educación superior empiezan a evidenciar un cambio de paradigma, de herramientas generativas aisladas a configuraciones más integradas. La presente revisión de alcance analizó la evidencia de los años 2024-2026, centrándose en las arquitecturas agénticas, feedback, personalización, funciones pedagógicas y supervisión humana. En la búsqueda que se realizó el 15 de agosto de 2026 en las bases de datos Scopus, Web of Science, ERIC, SciELO y Redalyc se encontraron 406 registros. Fueron incluidos 44 estudios para extracción y síntesis después del proceso de selección reportado mediante PRISMA-ScR. A partir de su lectura se reconocieron seis ejes, que posteriormente se organizaron en tres dimensiones: apoyo al aprendizaje, competencias e infraestructura, y condiciones de implementación. Entre los hallazgos, las arquitecturas agénticas mostraron una mayor explicitud técnica, mientras que la personalización y el feedback concentraron buena parte de las aplicaciones pedagógicas. Sin embargo, esa mayor complejidad técnica no estuvo acompañada de evidencia educativa con el mismo grado de desarrollo. Persisten vacíos en aprendizaje profundo, transferencia, equidad y evaluación longitudinal, lo que limita todavía la valoración de efectos educativos sostenidos.

Palabras Clave: inteligencia artificial agéntica; educación superior; feedback; RAG; sistemas multiagente; revisión de alcance.

1. Universidad Nacional Pedro Ruiz Gallo, Lambayeque – Perú
 2. Universidad Pontificia de San Francisco Xavier de Chuquisaca – Bolivia
 3. Benemérito Instituto Normal del Estado "Gral. Juan Crisóstomo Bonilla": Puebla, México
 4. Aden University – Panamá
 5. Universidad César Vallejo – Perú
 6. Universidad Nacional Jorge Basadre Grohman, Perú
- * Autor de correspondencia: jose.chafloque@unprg.edu.pe

ABSTRACT

Recent work on artificial intelligence in higher education suggests a shift from stand-alone generative tools toward more integrated configurations. This scoping review examined evidence published between 2024 and 2026, focusing on agentic architectures, feedback, personalization, pedagogical functions, and human oversight. Searches conducted on August 15, 2026, in Scopus, Web of Science, ERIC, SciELO, and Redalyc yielded 406 records. Following the selection process reported according to PRISMA-ScR, 44 studies were included for data extraction and synthesis. Six thematic axes were identified and subsequently organized into three dimensions: learning support, competencies and infrastructure, and implementation conditions. Agentic architectures showed greater technical explicitness, while personalization and feedback accounted for a substantial proportion of the pedagogical applications identified. However, this greater technical complexity was not accompanied by educational evidence at a comparable level of development. Gaps remain in deep learning, transfer, equity, and longitudinal evaluation, which continue to limit the assessment of sustained educational effects.

Keywords: agentic artificial intelligence; higher education; feedback; retrieval-augmented generation; multi-agent systems; scoping review.

INTRODUCCIÓN

La incorporación de inteligencia artificial en la educación superior se viene extendiendo desde herramientas generativas de propósito general hacia sistemas que combinan la recuperación de información, agentes especializados, analíticas de aprendizaje, reglas de decisión, memoria contextual y mecanismos de supervisión humana. En las configuraciones agénticas, determinadas tareas dejan de ser resueltas mediante una respuesta única y se distribuyen entre componentes que son capaces de consultar información, coordinar acciones, verificar resultados o ajustar su actuación conforme a objetivos definidos. Esta organización amplía las capacidades técnicas del sistema, aunque no supone por sí misma una mejora pedagógica. Desde una perspectiva educativa interesa que se examine qué funciones asumen estos componentes, cómo intervienen en el proceso de aprendizaje y qué decisiones permanecen bajo responsabilidad humana.

Se han identificado funciones que varían según el contexto de aplicación. Los estudios de Anghel et al. (2026) y Deng et al. (2026) incorporan agentes y mecanismos de verificación en procesos de evaluación automatizada. Modran (2025) y Xie et al. (2026) integran recuperación, orquestación y verificación dentro de la arquitectura del sistema. Bai et al. (2025), Belawati y Prasetyo (2025) y Qiu et al. (2026), en cambio, examinan las aplicaciones que están vinculadas con tutoría, asesoría y personalización. Estas diferencias dan a conocer que la arquitectura agéntica no depende solo de la denominación tecnológica, sino de la forma en que se distribuyen y coordinan las funciones dentro del sistema.

La expansión de estas aplicaciones tampoco equivale a una consolidación pedagógica. Algunos estudios informan sobre las mejoras en precisión, recuperación, participación, rendimiento o percepción de utilidad. Mientras que el aprendizaje profundo, la transferencia, la metacognición, la equidad y los efectos longitudinales aparecen con menor frecuencia entre los resultados evaluados. Fan et al. (2025), por ejemplo, encontraron una mejora del desempeño inmediato en escritura sin ganancias equivalentes en conocimiento o transferencia. Le et al. (2025) y Nendauni (2025), desde enfoques

diferentes, muestran que el feedback humano mantiene ese valor por dimensiones dialógicas, relacionales o contextuales que aún no son cubiertas por la automatización. Bozkurt et al. (2026) sitúan esta relación dentro de un sistema sociotécnico en el que la agencia humana continúa interviniendo en el uso y evaluación de la tecnología. Esta diversidad también dificulta la delimitación conceptual del campo. IA generativa, agentes basados en LLM, sistemas multiagente, RAG, Agentic RAG, sistemas tutores inteligentes, chatbots y aprendizaje adaptativo no describen necesariamente el mismo grado de autonomía ni la misma organización arquitectónica. A la vez, sistemas con denominaciones distintas pueden desempeñar funciones educativas similares. Por esta razón, la revisión consideró tanto la explicitud agéntica como la relación entre arquitectura, función pedagógica, feedback, personalización y supervisión humana.

En este contexto, la pregunta guía de esta revisión es: ¿cómo se configuran los sistemas de inteligencia artificial aplicados a la educación superior, especialmente las arquitecturas agénticas, en términos de funciones, roles, mecanismos de feedback, personalización, enfoques pedagógicos y supervisión humana, y qué vacíos presenta la evidencia publicada entre 2024 y 2026? El objetivo fue mapear la evidencia reciente sobre estos sistemas y aplicaciones, prestando especial atención a las arquitecturas agénticas y a sus roles funcionales, así como a los mecanismos de feedback, personalización, tutoría, evaluación, enfoques pedagógicos emergentes y formas de supervisión humana.

METODOLOGÍA

Se realizó una revisión de alcance (scoping review) orientada a mapear la extensión, naturaleza y organización de la evidencia reciente sobre inteligencia artificial aplicada a la educación superior, con énfasis en arquitecturas agénticas, mecanismos de feedback, personalización, tutoría, evaluación y supervisión humana. Se seleccionó este diseño porque el campo presenta heterogeneidad en tecnologías, funciones pedagógicas, diseños de estudio, poblaciones, métricas y niveles de validación. Por eso, el análisis no se planteó como una estimación única de efectividad. Interesaba, más bien, reconocer cómo se estaba organizando la evidencia y qué zonas permanecían menos desarrolladas

Arksey y O'Malley (2005) y Levac et al. (2010) sirvieron como referencias para ordenar las etapas de la revisión: pregunta, búsqueda, selección, extracción y síntesis. La lógica PCC se utilizó para delimitar la pregunta y los criterios de elegibilidad, mientras que el reporte siguió PRISMA-ScR (Peters et al., 2024; Tricco et al., 2018). La búsqueda se diseñó de manera amplia para captar tanto evidencia directamente agéntica como aplicaciones de IA educativa funcionalmente relacionadas; esta decisión permitió analizar la posición de las arquitecturas agénticas dentro de un campo tecnológico y pedagógico más amplio.

Tabla 1

Estructura PCC de la revisión.

Componente PCC	Definición operativa
Población	Educación superior: estudiantes universitarios, docentes, tutores, asesores, programas, cursos, plataformas o repositorios vinculados a contextos universitarios.
Concepto	Sistemas y aplicaciones de inteligencia artificial en educación superior con funciones pedagógicas, evaluativas o de apoyo al aprendizaje, con énfasis en arquitecturas agénticas, sistemas multiagente, RAG/Agentic RAG, tutoría inteligente, feedback automatizado, personalización, asesoría y mecanismos de supervisión humana.

Contexto	Educación superior presencial, virtual, híbrida o a distancia, incluyendo disciplinas como escritura académica, programación, ingeniería, odontología, medicina, ciencia de datos, inglés académico, formación investigativa y asesoría institucional.
----------	--

La búsqueda se realizó el 15 de agosto de 2026 en cinco bases: Scopus, Web of Science, ERIC, SciELO y Redalyc; seleccionadas para combinar cobertura multidisciplinaria internacional, literatura especializada en educación y producción regional iberoamericana. Las estrategias se construyeron mediante tres bloques conceptuales: (1) tecnologías de inteligencia artificial agéntica y generativa; (2) educación superior; y (3) funciones educativas relacionadas con feedback, evaluación, tutoría, personalización, escritura, alfabetización y autorregulación. Las cadenas se adaptaron a la sintaxis y capacidades de cada base. Los registros recuperados antes y después de los filtros se resumen en la Tabla 2; las estrategias completas, campos consultados, filtros, fechas y conteos detallados por fuente se presentan en la Tabla Suplementaria S1, siguiendo criterios de transparencia compatibles con PRISMA-S (Rethlefsen et al., 2021).

Tabla 2

Resultados de búsqueda por fuente antes y después de filtros.

Fuente	Registros iniciales	Registros tras filtros
Scopus	161	63
Web of Science	36	4
ERIC	130	25
SciELO	3	3
Redalyc	76	3
Total	406	98

Tabla 3

Criterios de elegibilidad aplicados.

Criterio	Inclusión	Exclusión
Periodo	Publicaciones entre 2024 y 2026.	Publicaciones fuera de 2024–2026. Los antecedentes metodológicos utilizados para fundamentar la revisión no forman parte del corpus.
Contexto	Estudios desarrollados en educación superior o dirigidos explícitamente a estudiantes, docentes, cursos, programas, asesoría académica o plataformas universitarias.	Estudios centrados en educación básica o secundaria sin transferencia explícita a educación superior.
Concepto tecnológico	Sistemas o aplicaciones de IA con al menos una de las siguientes características: IA agéntica, agentes LLM, sistemas multiagente, RAG/Agentic RAG, tutoría inteligente, feedback con IA, evaluación automatizada, aprendizaje adaptativo o IA generativa con función pedagógica explícita.	Herramientas digitales sin componente de IA, estudios de IA sin aplicación educativa o trabajos que no permiten identificar una función o aporte pertinente al mapeo.
Foco pedagógico	Relación explícita con tutoría, feedback, evaluación, personalización, asesoría, escritura o alfabetización académica, autorregulación, metacognición, generación	Estudios sin relación con aprendizaje, enseñanza, evaluación, apoyo académico, procesos formativos o

Tipo de evidencia	de recursos educativos o gobernanza/supervisión humana. Estudios empíricos, evaluaciones de sistemas, desarrollos o evaluaciones de prototipos, revisiones, análisis bibliométricos y marcos conceptuales o arquitectónicos que aportan información relevante para el mapeo.	condiciones de implementación educativa.
		Opinión, comentarios o textos sin base suficiente para extraer información pertinente al mapeo.
Disponibilidad	Texto completo recuperado para evaluación minuciosa de elegibilidad.	Texto completo no recuperado tras la búsqueda y la gestión de acceso; registro conservado únicamente para trazabilidad y no incorporado al corpus final.

Deduplicación, cribado y recuperación de textos completos

Los registros obtenidos después de aplicar los filtros definidos en cada base se consolidaron en una única matriz bibliográfica. La deduplicación se realizó mediante coincidencia de título, DOI, autores y metadatos bibliográficos, con revisión manual de posibles coincidencias. Se eliminaron 3 duplicados y el conjunto único quedó conformado por 95 registros. El primer cribado se realizó sobre título, resumen y palabras clave por dos revisores. Las discrepancias se discutieron hasta alcanzar consenso; los casos no conciliados se remitieron a un tercer revisor para decisión final. Sesenta y cinco registros requirieron consulta a texto completo. Cuando el título o el resumen no permitían resolver la elegibilidad, el estudio se mantenía en esta etapa y se intentaba obtener el documento antes de excluirlo. Para ello se recurrió, según disponibilidad, al acceso institucional, versiones de autor y fuentes abiertas. No todos pudieron recuperarse, lo que corresponde al 27,7% de los documentos buscados en esta etapa. Los registros sin texto completo quedaron documentados en la trazabilidad, pero no fueron evaluados en la fase final de elegibilidad. Dada la magnitud de esta pérdida, su posible efecto sobre la composición final del corpus se consideró posteriormente como una limitación relacionada con la disponibilidad de los textos.

Evaluación de elegibilidad a texto completo

Los 47 textos completos recuperados pasaron a evaluación de elegibilidad. Para cada registro se consideraron cuatro dominios definidos a priori: contexto de educación superior, presencia de una aplicación o sistema de IA pertinente, función educativa o condición de implementación relevante y suficiente información para contribuir al mapeo. Cada texto se revisó completo antes de asignar la decisión final. Cuando surgieron diferencias entre los dos revisores, el caso se discutió con base en esos dominios y, si no se alcanzaba acuerdo, intervenía un tercer revisor. El corpus final quedó integrado por 44 estudios. Otros 3 textos, aunque habían llegado a lectura completa, no reunieron de manera suficiente el foco conceptual o educativo previsto en el protocolo y fueron excluidos. Las diferencias entre los estudios incluidos no afectaron esa decisión de elegibilidad; se registraron después como características analíticas del corpus, entre ellas el tipo de evidencia, la explicitud arquitectónica y el eje temático.

Extracción, codificación y síntesis de datos

La extracción se organizó en la Tabla Suplementaria S2. Una parte de la matriz describía cada estudio (autor y año, tipo de evidencia, tecnología o arquitectura y contexto o población). Otra registraba la información que después sería utilizada en la síntesis: foco

y objetivo, rol funcional de la IA, mecanismo de feedback o evaluación, resultado principal, limitaciones, explicitud arquitectónica y eje temático.

Esa misma matriz se utilizó después para comprobar de qué estudios procedía cada categoría incluida en la síntesis. El tipo de evidencia y la explicitud arquitectónica se codificaron por separado, ya que un estudio empírico, técnico, secundario o conceptual podía describir sistemas con diferente grado de organización agéntica.

Para el tipo de evidencia se distinguieron estudios empíricos o intervencionales, evaluaciones técnicas o de prototipo, revisiones o síntesis y trabajos conceptuales o arquitectónicos. Los estudios secundarios se utilizaron para reconocer tendencias, problemas recurrentes y formas de organización del campo. Sus resultados no se interpretaron como evidencia independiente equivalente a la procedente de estudios primarios, debido a que algunas de las investigaciones sintetizadas en esas revisiones podían coincidir con trabajos primarios incluidos también en el corpus. Esta distinción se mantuvo durante la síntesis y la interpretación de los hallazgos. La explicitud arquitectónica se organizó en tres niveles. Nivel de arquitectura agéntica explícita, con agentes LLM, sistemas multiagente, Agentic RAG o mecanismos identificables de planificación, orquestación, verificación o escalamiento. Nivel de configuración híbrida o adaptativa que combinaba IA generativa o RAG con tutoría inteligente, aprendizaje adaptativo, reglas, memoria o aprendizaje por refuerzo, sin un sistema multiagente plenamente definido. Nivel de aplicación funcional de IA como chatbot, generador, tutor o mecanismo de feedback o evaluación, sin organización agéntica explícita. Cada estudio se le asignó a un eje principal según su aporte predominante; las relaciones con otras funciones se conservaron. Los seis ejes fueron: (1) tutoría, asesoría y personalización; (2) feedback y evaluación formativa; (3) alfabetización académica, escritura y metacognición; (4) arquitecturas agénticas, RAG y verificación; (5) gobernanza, ética y supervisión humana; y (6) síntesis del campo y tendencias.

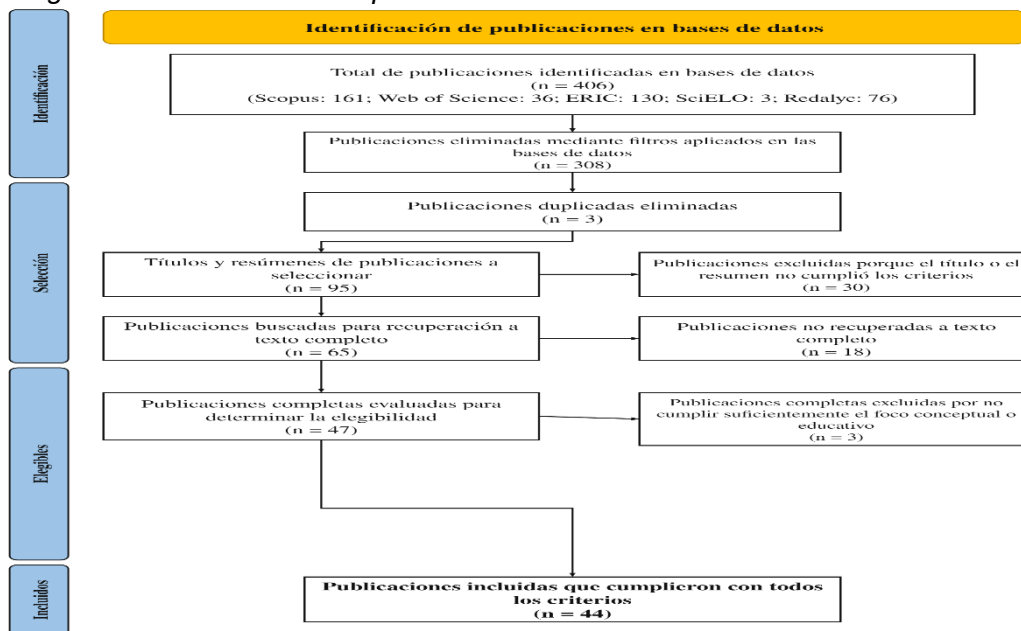
Los seis ejes se conservaron para describir el corpus. En la interpretación se leyeron además desde tres dimensiones: apoyo al aprendizaje (ejes 1–2), competencias e infraestructura (ejes 3–4) y condiciones de implementación (ejes 5–6). Esta segunda lectura permitió relacionar funciones distribuidas entre distintos ejes, sin modificar la clasificación de los estudios. Primero se realizó una descripción numérica de los estudios según los ejes temáticos y las características de la evidencia, cuyos resultados se presentan en las Tablas 4 y 5. Después se avanzó hacia una lectura temática del conjunto. En esta etapa interesó reconocer convergencias, diferencias y vacíos en la relación entre arquitectura, función pedagógica, feedback, personalización y supervisión humana. Debido a la heterogeneidad de diseños, tecnologías, muestras y métricas, no se realizó metaanálisis ni se utilizó una evaluación formal de riesgo de sesgo como criterio de exclusión. La interpretación de los hallazgos se mantuvo vinculada al tipo de evidencia disponible y no se formularon estimaciones globales de efectividad.

Flujo de selección y transparencia del reporte

El flujo de identificación, cribado, recuperación y selección se presenta en la Figura 1 mediante un diagrama PRISMA-ScR. La figura muestra únicamente las etapas de identificación, cribado, elegibilidad e inclusión; la distribución temática de los 44 estudios se presenta por separado en la Tabla 4.

Figura 1

Diagrama PRISMA-ScR del proceso de selección de estudios.



Nota: Diagrama de flujo adaptado de PRISMA 2020 y utilizado conforme a las recomendaciones de PRISMA-ScR para el reporte del proceso de selección en revisiones de alcance (Page et al., 2021; Tricco et al., 2018).

RESULTADOS

Distribución temática de los estudios

Los 44 estudios incluidos se distribuyeron en seis ejes temáticos definidos durante la codificación: tutoría, asesoría y personalización (n=14); feedback y evaluación formativa (n=11); alfabetización académica, escritura y metacognición (n=8); arquitecturas agénticas, RAG y verificación (n=6); gobernanza, ética y supervisión humana (n=3); y síntesis del campo y tendencias (n=2), como se resume en la Tabla 4. El detalle estudio por estudio se conserva en la Tabla Suplementaria S2.

Tabla 4

Distribución de estudios por eje de síntesis.

Eje de síntesis	N	Aporte al mapeo
Tutoría, asesoría y personalización	14	Sistemas de tutoría, asesoría académica, personalización, aprendizaje adaptativo y ajuste dinámico de tareas.
Feedback y evaluación formativa	11	Calificación, feedback automatizado, evaluación formativa, feedback informativo y aprendizaje autorregulado.
Alfabetización académica, escritura y metacognición	8	Escritura académica, literacidad de IA, preferencia de feedback, metacognición y agencia estudiantil.
Arquitecturas agénticas, RAG y verificación	6	RAG, Agentic RAG, multiagente, verificación factual, generación multimodal y pipelines de agentes.
Gobernanza, ética y supervisión humana	3	Human-over-the-loop, blockchain, auditoría, ética, transparencia y soberanía de datos.
Síntesis del campo y tendencias	2	Revisiones de campo, tendencias emergentes y desafíos transversales.

La composición del corpus según tipo de evidencia se presenta en la Tabla 5: 18 estudios empíricos o intervencionales (40,9%), 11 evaluaciones técnicas o de prototipo (25,0%), 11 revisiones o síntesis (25,0%) y 4 estudios conceptuales o arquitectónicos (9,1%). Esta distribución evidencia la coexistencia de resultados aplicados, validaciones técnicas y aportes secundarios o conceptuales dentro del mapa.

Tabla 5

Distribución del corpus por tipo de evidencia.

Tipo de evidencia	n	%
Empírica o intervencional	18	40,9
Evaluación técnica o de prototipo	11	25,0
Revisión o síntesis	11	25,0
Conceptual o arquitectónica	4	9,1

Tutoría, asesoría y personalización

El eje más numeroso reunió 14 estudios. Las propuestas de tutoría fueron diversas: incluyeron asistentes conversacionales, configuraciones adaptativas, sistemas institucionales y desarrollos parcialmente agénticos. Bai et al. (2025), mediante GPTutor, encontraron asociaciones con el compromiso conductual y emocional, pero no con el compromiso cognitivo. Belawati y Prasetyo (2025) reportaron, por su parte, mejoras significativas en participación y rendimiento, aunque los efectos observados fueron de magnitud reducida.

Los trabajos de López-Goyez et al. (2026a, 2026b) continúan esta línea de estudio con ELA Tutor, un sistema en Moodle que combina MAS, GenAI y aprendizaje por refuerzo. La adaptación de las estrategias tutoriales ocupa un lugar central en la propuesta, aunque los estudios mantienen la necesidad de supervisión docente y de validación en entornos reales. Yi y Wu (2025) presentan una distinta configuración: un único agente que detecta carga cognitiva y ajusta la dificultad de las tareas. Wang y Sun (2026) vinculan la tutoría basada en IA con engagement cognitivo, feedback docente y desempeño académico. En conjunto, estos trabajos muestran formas distintas de personalización, desde el ajuste de la dificultad o de estrategias hasta el seguimiento de variables relacionadas con el aprendizaje. En el ámbito de la asesoría académica, Arguson et al. (2026) evaluaron una arquitectura multiagente con aprendizaje por refuerzo y supervisión humana. El sistema redujo el tiempo medio de respuesta de 48,6 a 19,3 horas y mejoró la precisión de los planes, la detección de estudiantes en riesgo y la satisfacción de los usuarios. Tamascelli et al. (2025) encontraron modelos que utilizan una combinación de RAG, HITL, Gmail y búsqueda web en un chatbot que escala esa consulta cuando las fuentes institucionales no pueden resolverla. Qiu et al. (2026) evaluaron personalización explicable sobre un dataset, no se trató de una intervención educativa directa.

Feedback y evaluación formativa

En el eje de evaluación, la IA está relacionada a la calificación, verificación, feedback y apoyo en la autorregulación en relación con el trabajo realizado, aunque puede variar en la forma cómo se concreta en los distintos estudios. Así, en el modelo de Anghel et al. (2026) se integran mecanismos de verificación dentro de la evaluación que proporciona el modelo automatizado, en el de Deng et al. (2026) se apuesta por combinar agentes, LLM, SR-RAG y role-playing prompting para acercar sus resultados a los de la calificación humana. En el trabajo de Kundu et al. (2026) el enfoque es de otra configuración y se recurre a una solución de sistema multiagente que estructura hojas de respuestas

manuscritas antes de facilitar la calificación humana. En programación, Zhong y Zhan (2025) desplazan el foco hacia el efecto del feedback informativo, asociado con mejoras en el pensamiento computacional y en la motivación, junto con una disminución de la carga cognitiva.

Letteri y Vittorini (2025) reportan que la combinación de LLM con IA tradicional puede mejorar la percepción de utilidad y la comprensión del progreso en ciencia de datos, aunque también observaron un posible aumento de la carga cognitiva asociado con la verbosidad del feedback. Huang et al. (2026), desde un metaanálisis, identificaron un efecto positivo medio del feedback con IA sobre el aprendizaje autorregulado. Celik et al. (2024), por su parte, reconocen al feedback adaptativo entre las affordances más frecuentes de las herramientas de IA utilizadas en educación superior. Los resultados no apuntan exactamente a lo mismo. Mientras algunos trabajos muestran efectos favorables sobre aprendizaje o autorregulación, otros también dejan ver que la forma en que se presenta el feedback puede introducir nuevas dificultades. Las revisiones de Aguirre Tucto et al. (2026), Badger et al. (2026), Espino et al. (2026) y Larrea Eyzaguirre y Bustillos Torrez (2025) amplían el eje hacia otros contextos, entre ellos las competencias investigativas, el aprendizaje con chatbots y la educación médica u odontológica. Al tratarse de estudios secundarios, su aporte no es directamente comparable con el de las intervenciones o evaluaciones técnicas descritas antes. Estas revisiones permiten reconocer problemas recurrentes de mediación docente, transparencia, autorregulación y condiciones éticas de uso; la evaluación del proceso aparece con menor consistencia entre las fuentes.

Alfabetización académica, escritura y metacognición

En escritura académica, Cai (2026) y Pan et al. (2026) incorporan agentes o flujos instruccionales para apoyar revisión, verificación y tutoría. Chafloque-Capuñay et al. (2026) examinan esos usos junto con problemas de autoría, voz académica, evaluación del producto final y trazabilidad del proceso, además de funciones de planificación, revisión y feedback. La relación con el feedback humano introduce una tensión distinta. Le et al. (2025) encontraron que la preferencia por tutores humanos se mantuvo antes y después de la tarea, aunque la interacción libre con ChatGPT produjo algunos cambios en esa valoración. Nendauni (2025), desde otro contexto, muestra por qué esa preferencia puede persistir: los estudiantes valoraron el feedback humano por su carácter dialógico, relacional y sensible al contexto lingüístico. Fan et al. (2025) desplazan la cuestión hacia el aprendizaje. En su estudio, el uso de ChatGPT mejoró el desempeño en escritura, pero no produjo ganancias equivalentes en conocimiento o transferencia y estuvo acompañado de una menor presencia de procesos metacognitivos. La mejora inmediata, entonces, no parece avanzar necesariamente al mismo ritmo que el aprendizaje más profundo.

La alfabetización también aparece en experiencias que se alejan de la escritura académica formal. Tang et al. (2026) evaluaron GamifiedLM con estudiantes universitarios con dificultades leves de aprendizaje y reportaron resultados en motivación, foco, dominio de materiales y metacognición. Hu, Y.-H. et al. (2025) trabajaron con un agente pedagógico y un compañero virtual, asociados con mejoras en resolución de problemas de información, alfabetización informacional y necesidad de cognición. En estos trabajos, la alfabetización incluye también interpretar y verificar los aportes de la IA dentro del trabajo académico.

Arquitecturas agénticas, RAG y verificación

Seis estudios concentraron las arquitecturas agénticas más explícitas del corpus. Düsterbeck et al. (2026) trabajaron con Node-Sampling, un sistema multiagente orientado a mejorar la generación de preguntas de opción múltiple en educación médica. Modran

(2025) combinó RAG con Agent Communication Protocol y Model Context Protocol en una propuesta de tutoría adaptativa interoperable. Hu, Z. et al. (2025) aplicaron un sistema Agentic RAG local al análisis de repositorios GitHub en educación informática. Xie et al. (2026) desarrollaron SlideBot, donde recuperación, verificación factual y síntesis visual se integran dentro de un sistema multiagente para generar presentaciones académicas. La recuperación, la verificación y la distribución de tareas entre agentes especializados constituyeron rasgos recurrentes de este tipo de arquitecturas. Los componentes podían recuperar, evaluar, sintetizar, verificar, generar o informar, lo que las distingue de configuraciones que están únicamente basadas en RAG o prompting directos. La evidencia todavía procede de dominios restringidos, muestras pequeñas y sistemas sensibles a los costes computacionales y a la calidad de las fuentes, mientras que su relación con el aprendizaje ha sido estudiada con menor frecuencia.

Gobernanza, ética y supervisión humana

Tres estudios fueron fundamentalmente abordados a partir del eje de la gobernanza, a pesar de que se puede encontrar dicha preocupación también en otras partes del corpus. Klepikov et al. (2026) incorporan auditoría, roles especializados y control humano dentro de una arquitectura diagnóstica basada en LLM. Hattab y Ziadi (2026) entrelazan sistemas multiagentes y la blockchain para introducir trazabilidad y supervisión de las propias evaluaciones de la informática. Bozkurt et al. (2026) intentan responder a dicho problema desde un enfoque sociotécnico relacionado con la tecnología educativa, la agencia humana, la ética y las condiciones institucionales de implementación. La supervisión humana tiene la función de controlar, validar y escalar cuando el sistema no puede resolver una tarea de manera autónoma. Klepikov et al. (2026) incluyen un principio de Human-over-the-Loop dentro de su arquitectura diagnóstica y Tamascelli et al. (2025) mantienen los mecanismos HITL para tareas que requieren intervención humana. Estas formas de supervisión humana aparecen en los sistemas con mayor autonomía técnica, especialmente para tareas con consecuencias evaluativas o académicas.

Síntesis del campo y tendencias

Rahman et al. (2026) describen un campo en expansión donde hay un mayor uso de tutoría inteligente, de evaluación adaptativa, de simulaciones inmersivas y, más recientemente, de feedback basado en LLM e IA generativa. Los resultados que muestran son favorables en la relación al desempeño y la participación, aunque con menos claridad en el ámbito de la ética, equidad y efectos a largo plazo. En cuanto al trabajo de Simarco-Scarci et al. (2026) se observa ese mismo crecimiento desde el ámbito de la educación superior, aunque la atención se desplaza hacia las condiciones de implementación: orientación al estudiantado, regulación del uso, seguridad de los datos y control del sesgo algorítmico. En definitiva, ambas revisiones muestran que la expansión de aplicaciones avanza más rápido que la consolidación de criterios para su adopción y evaluación sostenida.

DISCUSIÓN

La evidencia revisada muestra que la IA agéntica no constituye todavía un espacio completamente separado del resto de la IA educativa. Varias funciones ya presentes en tutoría, evaluación o personalización empiezan a distribuirse entre componentes especializados, con recuperación, verificación, coordinación y supervisión dentro de la propia arquitectura. Ese cambio es visible en los trabajos de Anghel et al. (2026), Deng et al. (2026), Modran (2025), Hu, Z. et al. (2025) y Xie et al. (2026), donde mejoran métricas

de acuerdo, recuperación, fidelidad o calidad de los artefactos. La dificultad aparece cuando esas mejoras se trasladan al plano educativo, porque la mayoría de estos resultados siguen vinculados con tareas y dominios concretos.

En personalización y feedback, la evidencia incluida no sigue una dirección uniforme. Los estudios muestran efectos favorables en dimensiones como participación, rendimiento, percepción de utilidad, compromiso y aprendizaje autorregulado, pero esos resultados corresponden a medidas y diseños diferentes y no permiten asumir un efecto pedagógico equivalente entre intervenciones (Bai et al., 2025; Belawati & Prasetyo, 2025; Letteri & Vittorini, 2025; Huang et al., 2026). Esta heterogeneidad también aparece en la forma de la asistencia: una mayor disponibilidad de feedback puede facilitar la comprensión del progreso, mientras que su extensión o complejidad puede introducir carga cognitiva adicional. Por ello, el problema no parece residir únicamente en disponer de feedback generado por IA, sino en cómo se diseña, en qué momento interviene y qué resultado educativo se utiliza para evaluarlo.

Una segunda diferencia aparece cuando se distingue entre desempeño inmediato y aprendizaje. La evidencia del corpus muestra que una mejora en la ejecución de una tarea no necesariamente se acompaña de ganancias equivalentes en conocimiento, transferencia o actividad metacognitiva. A la vez, la preferencia por el feedback humano persiste cuando los estudiantes valoran dimensiones dialógicas, relacionales o sensibles al contexto que la asistencia automatizada no reproduce del mismo modo (Fan et al., 2025; Le et al., 2025; Nendauni, 2025). Estas diferencias sugieren que el producto obtenido y el proceso de aprendizaje deben analizarse como resultados relacionados, pero no intercambiables.

Esta diferencia también depende de la composición del corpus. Algunos estudios evalúan precisión, recuperación, fidelidad de citas, acuerdo con evaluadores o calidad de los artefactos, mientras que otros examinan aprendizaje, transferencia, pensamiento crítico, metacognición o retención. Estos resultados responden a preguntas distintas y no deben interpretarse en un mismo nivel. En particular, la evidencia sobre efectos educativos sostenidos es más limitada y procede de diseños, periodos de seguimiento y medidas heterogéneas. Por ello, la mayor disponibilidad de resultados técnicos no permite inferir una mejora equivalente en los procesos de aprendizaje.

La supervisión humana tampoco llega a cumplir una función uniforme. Klepikov et al. (2026) incorporan control y auditoría dentro de la arquitectura; Tamascelli et al. (2025) mantienen mecanismos HITL para situaciones que requieren intervención; Hattab y Ziadi (2026) enfatizan la trazabilidad, mientras que Bozkurt et al. (2026) sitúan la cuestión en un marco sociotécnico. La intervención humana puede formar parte del funcionamiento operativo del sistema o actuar como mecanismo de control y responsabilidad institucional. Distinción que resulta muy pertinente en la evaluación, el diagnóstico y la asesoría, donde debe establecerse quién valida las decisiones y en qué condiciones corresponde escalar la intervención a una persona. En el corpus aparecen de manera recurrente arquitectura verificable, feedback formativo, mediación humana, alfabetización crítica y validación empírica. No forman todavía un modelo pedagógico ni una secuencia causal. Su relación cambia según la aplicación y el tipo de evidencia disponible. Rahman et al. (2026) y Simarco-Scarci et al. (2026) sitúan estas tensiones dentro de un campo que sigue ampliando aplicaciones mientras mantiene vacíos en ética, equidad, regulación, seguridad de los datos y evaluación de largo plazo.

Durante la recuperación de los documentos apareció una primera dificultad que afecta la lectura del corpus final. De los 65 registros que llegaron a la etapa de búsqueda del texto completo, 18 no pudieron ser recuperados. Esto equivale al 27,7 % de los documentos que se intentó examinar en esa fase. Las razones no fueron las mismas en todos los casos:

hubo restricciones de acceso, documentos que no estaban disponibles por las vías institucionales consultadas y algunos casos en los que tampoco se obtuvo respuesta de los autores. Estos registros se mantuvieron dentro de la trazabilidad de la revisión, pero no fue posible evaluar su elegibilidad. Por ello, no puede saberse si su incorporación habría modificado parcialmente la composición del corpus o alguno de los patrones identificados.

La búsqueda también se delimitó a Scopus, Web of Science, ERIC, SciELO y Redalyc y al periodo 2024–2026. Esta decisión permitió concentrar la revisión en la producción más reciente asociada con IA generativa y configuraciones agénticas, aunque deja fuera desarrollos anteriores vinculados con sistemas tutores inteligentes, agentes pedagógicos o aprendizaje adaptativo. Algunos de esos antecedentes podrían ayudar a entender mejor la continuidad entre tecnologías previas y las arquitecturas que actualmente se describen como agénticas.

Otra dificultad provino de la propia composición del corpus. Los estudios incluidos no responden a un mismo tipo de diseño ni producen el mismo tipo de evidencia. Se encontraron investigaciones empíricas, evaluaciones técnicas, prototipos, revisiones, análisis bibliométricos y trabajos conceptuales o arquitectónicos. Esa diversidad era esperable en una revisión de alcance, pero obliga a tener cuidado al comparar sus resultados. Una mejora en precisión, recuperación o verificación dentro de un sistema no tiene el mismo alcance que un resultado observado en aprendizaje, transferencia o metacognición.

Algo similar ocurrió con el lenguaje utilizado para describir las tecnologías. Términos como IA generativa, agentes LLM, sistemas multiagente, RAG, Agentic RAG, sistemas tutores inteligentes y aprendizaje adaptativo aparecieron con cierta proximidad entre estudios, aunque no siempre describían el mismo nivel de autonomía ni una arquitectura equivalente. La clasificación por función predominante y explicitud arquitectónica ayudó a mantener esas diferencias durante el análisis, pero no elimina completamente una ambigüedad que todavía pertenece al propio campo.

También fue necesario distinguir el lugar que ocupaban los estudios secundarios. Algunas revisiones o síntesis podían incorporar investigaciones primarias que, a su vez, estaban presentes de manera independiente en el corpus. Por esa razón, estos trabajos se utilizaron principalmente para reconocer tendencias, problemas recurrentes y formas de organización del campo, y no como evidencia adicional equivalente a la procedente de estudios primarios.

Finalmente, la profundidad de la evidencia fue desigual. En una parte de los estudios predominaban simulaciones, muestras pequeñas, dominios específicos, percepciones de usuarios o métricas técnicas. Las evaluaciones prolongadas de aprendizaje profundo, transferencia, pensamiento crítico, equidad o retención aparecieron con menor frecuencia. Esta diferencia condiciona la interpretación de los hallazgos: el corpus permite describir con mayor claridad lo que estas arquitecturas hacen técnicamente que establecer, con el mismo nivel de evidencia, qué efectos educativos mantienen en el tiempo.

CONCLUSIONES

En el análisis, la explicitud agéntica es más visible cuando funciones como recuperación, verificación, adaptación, coordinación y supervisión dejan de ser un único componente y pasan a distribuirse dentro de la arquitectura. Estas configuraciones no sustituyen completamente los sistemas generativos ni a los tutores inteligentes menos explícitamente agénticos, sino que conviven con ellos. La diferencia técnica puede reconocerse con mayor claridad cuando su alcance sobre el aprendizaje sigue siendo menos definido.

La evidencia es más consistente al presentar mejoras en recuperación, precisión, verificación y generación que al informar cambios en el aprendizaje profundo, transferencia o metacognición. Esa diferencia también aparece en personalización y feedback, donde disponer de más apoyo no significa necesariamente que cambien, en la misma medida, los procesos de aprendizaje. En esa relación continúan interviniendo la agencia del estudiante y la mediación humana.

Con la evidencia disponible resulta prematuro que se pueda hablar de un modelo pedagógico consolidado alrededor de la IA agéntica. Supervisión humana, verificabilidad, feedback formativo, alfabetización crítica y validación empírica aparecen de forma recurrente en el corpus, pero no queda suficientemente establecido cómo se articulan entre sí ni qué peso tienen sobre el aprendizaje. Sigue abierta, por tanto, una cuestión central: en qué momento una mayor complejidad arquitectónica deja de representar solo una mejora técnica y empieza a producir cambios educativos que puedan sostenerse en el tiempo.

REFERENCIAS

- Aguirre Tucto, J. S., Crespo Barrionuevo, K. B., Bernuy Pimentel, J. J., & Rios García, G. (2026). Componentes teóricos y metodológicos de la IA generativa en competencias investigativas universitarias: Revisión sistemática. *Revista Aula Virtual*, 7(15), 1–23. <https://doi.org/10.5281/zenodo.21309447>
- Anghel, C., Anghel, A. A., Craciun, M. V., Cocu, A., Vulpe, D.-E., Andrei, C. A., Maier, C., Scheau, C., Dragosloveanu, S., & Cergan, R. (2026). GradeAgentOps: A verification-first framework for evidence-anchored LLM exam grading. *AI*, 7(6), Article 198. <https://doi.org/10.3390/ai7060198>
- Arguson, A. C., Sabonsolin, J. A., Malasaga, E. V., & Ramos, R. F. (2026). A hybrid framework for LMS web analytics: Combining clickstream analysis and sentiment mining for retention prediction. In *Proceedings of the 9th International Conference on Computers in Management and Business (ICCMB 2026)*, pp. 173–179. Association for Computing Machinery. <https://doi.org/10.1145/3802463.3802490>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Badger, M., López, N., & Nissen, C. F. R. (2026). The impact of GenAI chatbots on student learning in higher education: A literature review. *International Journal of Technology in Education*, 9(1), 43–69. <https://doi.org/10.46328/ijte.5111>
- Bai, H., Lui, W. C., & Khiatani, P. V. (2025). Promoting student engagement with GPTutor: An intelligent tutoring system powered by generative AI. *International Journal of Educational Technology in Higher Education*, 22, Article 77. <https://doi.org/10.1186/s41239-025-00571-9>
- Belawati, T., & Prasetyo, D. (2025). Generative AI-based tutoring for enhancing learning engagement and achievement. *Open Praxis*, 17(2), 211–226. <https://doi.org/10.55982/openpraxis.17.2.902>
- Bozkurt, A., Crompton, H., Farrow, R., Kukulska-Hulme, A., Dron, J., West, R., Palalas, A., Bower, M., Xiao, J., Tlili, A., Henriksen, D., Pazurek, A., Huijser, H., Chiu, T. K. F., Jandrić, P., Jordan, K., Curry, J., Kimmons, R., Cukurova, M., Reeves, T., Hwang, G.-J., Shea, P., Lodge, J., Weller, M., Ng, D., & Asino, T. I. (2026). Redefining educational technology: A critical collaborative inquiry. *Open Praxis*, 18(2), 192–211. <https://doi.org/10.55982/openpraxis.18.2.1117>

- Cai, Y. (2026). A multi-agent collaborative system for student writing assistance: An empirical study on mitigating AI bias and hallucination. In *Proceedings of the 2026 International Conference on Big Data and Informatization Education (ICBDIE '26)* (pp. 314–321). Association for Computing Machinery. <https://doi.org/10.1145/3806980.3807031>
- Celik, I., Gedrimiene, E., Siklander, S., & Muukkonen, H. (2024). The affordances of artificial intelligence-based tools for supporting 21st-century skills: A systematic review of empirical research in higher education. *Australasian Journal of Educational Technology*, 40(3), 19–38. <https://doi.org/10.14742/ajet.9609>
- Chafloque-Capuñay, J. E., Uría-Sánchez, M. B., Gonzales-Velásquez, E. A., Borges-Huanca, A., Marocho-Rojas, A., & Ussa-Lopez, Y. E. (2026). Inteligencia artificial generativa y escritura académica en la educación superior: Revisión integrativa de sus implicaciones pedagógicas, éticas y evaluativas. *Prohominum*, 8(2), 247–262. <https://doi.org/10.47606/ACVEN/PH0487>
- Deng, X., Miletic, V., Trinh, E., Gao, J., Xu, C., & Liu, D. (2026). DentEval: Fine-tuning-free expert-aligned assessment in dental education via LLM agents. In J. C. Gee, J. Hong, C. H. Sudre, P. Golland, D. C. Alexander, J. E. Iglesias, A. Venkataraman, & J. H. Kim (Eds.), *Medical Image Computing and Computer Assisted Intervention, MICCAI 2025 - 28th International Conference, 2025, Proceedings* (pp. 144–153). Springer. https://doi.org/10.1007/978-3-032-04971-1_14
- Düsterbeck, L. M., Größler, M., Credidio, G., Bellmann, L., & Riemann, L. T. (2026). Node-Sampling: Adaptive multi-agent optimization in medical education. *Frontiers in Computer Science*, 8, Article 1758333. <https://doi.org/10.3389/fcomp.2026.1758333>
- Espino, L. C., Antonio, R. P., Constantino, W. R., Espino, C. L., & Wider, W. (2026). Artificial intelligence-driven teaching methods for enhancing higher quality education: A bibliometric analysis. *International Journal of Technology in Education*, 9(1), 153–167. <https://doi.org/10.46328/ijte.5114>
- 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(2), 489–530. <https://doi.org/10.1111/bjet.13544>
- Hattab, S., & Ziadi, F. (2026). Towards secure and transparent practical assessments in computer science: A multi-agent and blockchain-based architecture. In *Proceedings of the 18th International Conference on Agents and Artificial Intelligence (ICAART 2026)* (Vol. 1, pp. 504–511). SCITEPRESS. <https://doi.org/10.5220/0014412600004052>
- Hu, Y.-H., Yu, H.-Y., & Hsieh, C.-L. (2025). Can pedagogical agent-based scaffolding boost information problem-solving in one-on-one collaborative learning with a virtual learning companion? *Education and Information Technologies*. Advance online publication. <https://doi.org/10.1007/s10639-025-13784-2>
- Hu, Z., Paprotskyi, M.-M., Vysotska, V., Chyrun, L., Ushenko, Y., & Uhryn, D. (2025). Local agentic RAG-based information system development for intelligent analysis of GitHub code repositories in computer science education. *International Journal of Modern Education and Computer Science*, 17(5), 109–145. <https://doi.org/10.5815/ijmecs.2025.05.07>
- Huang, C.-Q., Lu, L.-N., Huang, Q.-H., Zhang, Y.-R., He, T., Tu, Y.-F., & Hwang, G.-J. (2026). Effects of artificial intelligence feedback on students' self-regulated learning in higher education: A three-level meta-analysis. *Educational Psychology Review*. Advance online publication. <https://doi.org/10.1007/s10648-026-10166-z>

- Klepikov, O. E., Pozdnyakova, E. M., & Sizov, Y. O. (2026). From measurement to understanding: An architecture of emulated expertise for LLM-diagnostics in education. *Psychology. Journal of the Higher School of Economics*, 23(2), 263–274. <https://doi.org/10.17323/1813-8918-2026-2-263-274>
- Kundu, S., Chatterjee, A., Dhokne, N., Goyal, R. K., & Mishra, S. (2026). AgiQ-A: An agentic question-answer pairing system to aid handwritten answer sheet evaluation. In *Proceedings of the 18th International Conference on Computer Supported Education (CSEDU 2026)* (Vol. 1, pp. 94–105). SCITEPRESS. <https://doi.org/10.5220/0014820600004021>
- Larrea Eyzaguirre, J. C., & Bustillos Torrez, W. (2025). La inteligencia artificial generativa en la educación odontológica: Una revisión narrativa de aplicaciones y desafíos. *Revista de la Asociación Odontológica Argentina*, 113(3), Artículo e1131252. <https://doi.org/10.52979/raoa.1131252.1306>
- Le, H., Shen, Y., Li, Z., Xia, M., Tang, L., Li, X., Jia, J., Wang, Q., Gašević, D., & Fan, Y. (2025). Breaking human dominance: Investigating learners' preferences for learning feedback from generative AI and human tutors. *British Journal of Educational Technology*, 56(5), 1758–1783. <https://doi.org/10.1111/bjet.13614>
- Letteri, I., & Vittorini, P. (2025). Enhancing student feedback in data science education: Harnessing the power of AI-generated approaches. *International Journal of Artificial Intelligence in Education*, 35, 3071–3094. <https://doi.org/10.1007/s40593-025-00492-8>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5, 69. <https://doi.org/10.1186/1748-5908-5-69>
- López-Goyez, J. P., González-Briones, A., & Demazeau, Y. (2026a). An adaptive multi-agent architecture with reinforcement learning and generative AI for intelligent tutoring systems: A Moodle-based case study. *Applied Sciences*, 16(3), Article 1323. <https://doi.org/10.3390/app16031323>
- López-Goyez, J. P., González-Briones, A., Demazeau, Y., & Guaytarilla G., J. M. (2026b). An exploratory study of a generative AI-based intelligent tutoring system using a multi-agent architecture in higher education. *Applied Sciences*, 16(11), Article 5453. <https://doi.org/10.3390/app16115453>
- Modran, H. A. (2025). Leveraging RAG with ACP & MCP for adaptive intelligent tutoring. *Applied Sciences*, 15(21), Article 11443. <https://doi.org/10.3390/app152111443>
- Nendauni, L. (2025). Whose feedback matters? Exploring human and AI-supported writing feedback practices in a South African writing centre. *Journal of Learning Development in Higher Education*, 38. <https://doi.org/10.47408/jldhe.vi38>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pan, J., Hu, Y., & Chen, H. (2026). A technology-enhanced learning framework for college English writing: Orchestrating generative AI agents in process-based instruction. In *Proceedings of the 3rd International Conference on Informatics Education and Computer Technology Applications (IECA 2026)*, pp. 930–937. Association for Computing Machinery. <https://doi.org/10.1145/3802133.3802283>
- Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2024). Scoping reviews. In E. Aromataris, C. Lockwood, K. Porritt, B. Pilla, & Z. Jordan (Eds.), *JBIM Manual for Evidence Synthesis*. JBI. <https://doi.org/10.46658/JBIMES-24-09>

- Qiu, C., Cui, K., & Wang, Z. (2026). Personalized driven instruction through explainable agentic AI in multicultural higher education environments. *Big Data*. Advance online publication. <https://doi.org/10.1177/2167647X261463927>
- Rahman, A., Khandakar, A., Ayari, M. A., Naji, K., Al-Ali, A. K., Sellami, A. L., & Alhazbi, S. M. A. (2026). Artificial intelligence innovations, challenges and emerging trends in engineering education. *Discover Education*, 5, Artículo 179. <https://doi.org/10.1007/s44217-026-01137-1>
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10, Article 39. <https://doi.org/10.1186/s13643-020-01542-z>
- Simarco-Scarci, A., Fonseca, M. H., Moser-Teixeira, T., & Fleig-Dal Forno, L. (2026). Gestión del conocimiento e inteligencia artificial en la educación superior. *Alteridad*, 21(1), 108–122. <https://doi.org/10.17163/alt.v21n1.2026.08>
- Tamascelli, M., Bunch, O., Fowler, B., Taeb, M., & Cohen, A. (2025). Academic advising chatbot powered with AI agent. *Proceedings of the 2025 ACM Southeast Conference*, 195–202. <https://doi.org/10.1145/3696673.3723065>
- Tang, K., Chen, K., Gopi, S., Xie, L., Quinlan, M., & Cho, Y. (2026). GamifiedLM: Co-designing an LLM-driven gamified learning app with university students to mitigate learning difficulties. In *Proceedings of the 2026 Designing Interactive Systems Conference* (DIS '26, pp. 2450–2473). ACM. <https://doi.org/10.1145/3800645.3812916>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Wang, G., & Sun, F. (2026). Effect of artificial intelligence-based tutoring, cognitive engagement, and teacher feedback on university students' academic performance. *European Journal of Education*, 61(2), Article e70619. <https://doi.org/10.1111/ejed.70619>
- Xie, Y., Fan, Z., Pei, J., Sun, Y., & Chang, K.-W. (2026). SlideBot: A multi-agent framework for generating informative, reliable, multi-modal presentations. *Proceedings of the 34th ACM International Conference on Information and Knowledge Management* (CIKM '26, pp. 1–10). Association for Computing Machinery. <https://doi.org/10.1145/3789876.3789901>
- Yi, Y., & Wu, X. (2025). Cognitive load early warning based on behavioral data: A single-agent approach for dynamically adjusting task difficulty. In *Proceedings of the 2nd International Conference on Big Data Analytics and Artificial Intelligence Application* (BDAIA 2025, pp. 140–144). Association for Computing Machinery. <https://doi.org/10.1145/3788108.3788518>
- Zhong, X., & Zhan, Z. (2025). An intelligent tutoring system for programming education based on informative tutoring feedback: System development, algorithm design, and empirical study. *Interactive Technology and Smart Education*, 22(1), 3–24. <https://doi.org/10.1108/ITSE-09-2023-0182>