



Doi: <https://doi.org/10.70577/asce.v5i2.844>

Recibido: 2026-04-21

Aceptado: 2026-05-08

Publicado: 2026-06-20

Inteligencia Artificial Generativa y Educación Matemática: Tensiones entre el uso Instrumental y el Aprendizaje Complejo

Generative Artificial Intelligence and Mathematics Education: Tensions Between Instrumental Use and Complex Learning

Autores

Melvin Ramírez Bogantes ¹

meramirez@itcr.ac.cr

<https://orcid.org/0000-0001-5516-0085>

Tecnológico de Costa Rica

Heredia - Costa Rica

Kenneth Mejía Luna ²

kmluna16@estudiantec.cr

<https://orcid.org/0009-0001-8485-995X>

Tecnológico de Costa Rica

Cartago - Costa Rica

Como Citar

Ramírez Bogantes. M. &, Mejía Luna. K. (2026) Inteligencia Artificial Generativa y Educación Matemática: Tensiones entre el uso Instrumental y el Aprendizaje Complejo ASCE MAGAZINE 5(2) 1605-1627



Resumen

La incorporación acelerada de la inteligencia artificial generativa (IAGen) en la educación matemática plantea interrogantes sobre su capacidad para promover aprendizajes cognitivamente complejos. La literatura crece de forma exponencial desde la liberación de los modelos conversacionales, pero buena parte del campo se orienta hacia la eficiencia instruccional, la adopción tecnológica y la personalización, en detrimento de los procesos de alta demanda cognitiva. El objetivo fue mapear y analizar la presencia de tres dimensiones cognitivas complejas, modelización matemática, autorregulación del aprendizaje y pensamiento crítico, en la literatura sobre IAGen y educación matemática publicada entre 2020 y 2025. Se adoptó un enfoque cualitativo con diseño exploratorio y descriptivo, mediante una revisión de alcance bajo el protocolo PRISMA-ScR; se recuperaron 219 artículos en seis bases de datos (Web of Science, Scopus, ERIC, SciELO, IEEE Xplore y ACM Digital Library) y se codificaron 34 variables por estudio. Los resultados muestran una aceleración exponencial de la producción (51,6 % en 2025), dominada por revisiones sistemáticas (80,4 %) y por ChatGPT/GPT (63,9 %), con la resolución de problemas en el 77,2 % de los procesos matemáticos abordados. La autorregulación aparece en el 31,5 % del corpus, la modelización en el 13,2 % y el pensamiento crítico en el 8,2 %; el 58,4 % no aborda ninguna y solo el 2,3 % las integra simultáneamente. Se concluye que el corpus revela cinco tensiones estructurales entre uso instrumental y aprendizaje complejo, y cuatro vacíos que configuran una agenda prioritaria de investigación empírica longitudinal en el contexto iberoamericano.

Palabras clave: inteligencia artificial generativa, educación matemática, revisión de alcance, modelización matemática, autorregulación del aprendizaje, pensamiento crítico.



Abstract

The accelerated incorporation of generative artificial intelligence (GenAI) into mathematics education raises questions about its capacity to foster cognitively complex learning. The literature has grown exponentially since the release of conversational models, yet a large portion of the field remains oriented toward instructional efficiency, technology adoption, and personalization, at the expense of high cognitive demand processes. The objective was to map and analyze the presence of three complex cognitive dimensions, mathematical modeling, self-regulated learning, and critical thinking, in the literature on GenAI and mathematics education published between 2020 and 2025. A qualitative approach with an exploratory and descriptive design was adopted, through a scoping review following the PRISMA-ScR protocol; 219 articles were retrieved from six databases (Web of Science, Scopus, ERIC, SciELO, IEEE Xplore, and ACM Digital Library) and 34 variables were coded per study. The results show an exponential acceleration in production (51.6 % in 2025), dominated by systematic reviews (80.4 %) and by ChatGPT/GPT (63.9 %), with problem solving accounting for 77.2 % of the mathematical processes addressed. Self-regulated learning appears in 31.5 % of the corpus, mathematical modeling in 13.2 %, and critical thinking in 8.2 %; 58.4 % address none and only 2.3 % integrate them simultaneously. It is concluded that the corpus reveals five structural tensions between instrumental use and complex learning, and four research gaps that shape a priority agenda for longitudinal empirical research situated in the Ibero-American context.

Keywords: generative artificial intelligence, mathematics education, scoping review, mathematical modeling, self-regulated learning, critical thinking.

Introducción

La irrupción de la Inteligencia Artificial Generativa (IAGen), en particular ChatGPT desde noviembre de 2022, ha modificado el debate educativo global de forma acelerada. En el dominio de la educación matemática, herramientas capaces de resolver integrales, generar problemas, tutorizar paso a paso y proporcionar retroalimentación conversacional instalan una pregunta que los marcos disponibles no responden con suficiente precisión: ¿qué queda del aprendizaje matemático cuando la herramienta puede completar la tarea por el estudiante? Esta pregunta no es retórica: define la tensión constitutiva del campo en el período 2020 a 2025 y orienta el presente estudio.

Revisiones previas documentan el crecimiento exponencial de la producción sobre Inteligencia Artificial (IA) en educación matemática desde 2022. Nguyen y Pham (2025), en un corpus de 58 artículos para el período 2015 a 2025, identificaron una concentración en tutoría adaptativa y sistemas tutores inteligentes (ITS, por sus siglas en inglés), con escasa atención a procesos de alta demanda cognitiva. Panqueban y Huincahue (2024), en una revisión de 29 artículos, constataron la hegemonía de marcos tecnológicos sobre marcos propios de la didáctica matemática. Yi et al. (2025) realizaron un metaanálisis centrado en la efectividad de la IA sobre el rendimiento matemático en K-12 (educación primaria y secundaria), con énfasis en variables de desempeño académico antes que en el desarrollo de competencias de alta complejidad cognitiva. Pepin et al. (2025), por su parte, documentaron un crecimiento explosivo de la literatura sobre ChatGPT en educación matemática y advirtieron sobre la falta de fundamentación pedagógica que acompaña esa expansión. Opesemowo y Adewuyi (2024) confirmaron, en una revisión sistemática del campo, que la integración de IA en educación matemática se ha producido en el marco de la cuarta revolución industrial pero con preferencia por enfoques instrumentales antes que por enfoques didácticos fundamentados. Ninguna de estas revisiones, sin embargo, cuantificó sistemáticamente la presencia o ausencia de las tres dimensiones cognitivas de mayor complejidad formativa (modelización, autorregulación y pensamiento crítico) en un corpus amplio para el período más reciente.

A pesar de ese crecimiento, persiste una problemática estructural. La literatura del campo continúa orientándose hacia la adopción instrumental de las herramientas, la eficiencia instruccional y la personalización adaptativa, mientras que las competencias

cognitivamente complejas que la educación matemática contemporánea demanda permanecen marginales en la investigación. La modelización matemática es la competencia de mayor transferencia al mundo extraescolar (Borromeo Ferri, 2021; Greefrath y Vorhölter, 2022). La autorregulación del aprendizaje constituye la condición psicológica necesaria para que la retroalimentación de la IA se convierta en aprendizaje y no en sustitución cognitiva (Contel y Cusi, 2025; Panadero y Lipnevich, 2022). El pensamiento crítico frente a las respuestas producidas por los modelos grandes de lenguaje (Large Language Model, LLM) es la competencia cuya ausencia hace más vulnerable al estudiante ante los errores sistemáticos y las demostraciones plausibles pero falsas que estos sistemas pueden generar (Urhan et al., 2024; Yoon et al., 2024). Documentar la presencia o ausencia de estas tres dimensiones permite trazar la brecha entre lo que el campo investiga y lo que los currículos contemporáneos demandan.

La pertinencia del estudio se acentúa en el contexto iberoamericano. Chacón-Rivadeneira et al. (2024) documentaron que la adopción de IA en la educación matemática latinoamericana enfrenta desafíos estructurales relacionados con la infraestructura, la formación docente y las políticas públicas; estos desafíos son cualitativamente distintos a los reportados en contextos anglosajones. En el caso costarricense, Campos (2024) evidenció la urgencia de directrices pedagógicas y éticas claras que regulen la adopción tecnológica en educación superior, mientras que Acuña y Marrero (2025) demostraron que la integración de IA en cursos universitarios de Cálculo mediante metodologías activas es viable, pero su éxito depende de una fundamentación teórica que trascienda el uso instrumental. Ramírez-Bogantes et al. (2026) y Meza-Cascante et al. (2026) profundizaron en esta línea para el caso del Tecnológico de Costa Rica, evidenciando que la innovación didáctica con IAGen exige posicionarla como mediadora cognitiva y ética, y no meramente como herramienta de productividad.

A partir de este diagnóstico, se planteó como objetivo general mapear y analizar la literatura publicada entre 2020 y diciembre de 2025 sobre IA en educación matemática, con el fin de identificar tendencias bibliométricas y temáticas, la presencia de dimensiones cognitivas complejas y los vacíos que configuran la frontera actual del campo. Las preguntas que orientan el estudio son: (a) ¿qué características bibliométricas presenta el corpus?, (b) ¿cuáles son las tendencias temáticas dominantes en los artículos

analizados?, (c) ¿qué lugar ocupan la modelización, la autorregulación y el pensamiento crítico en ese corpus?, y (d) ¿qué vacíos y tensiones sugiere la literatura analizada? La tesis que este estudio desarrolla es que el campo presenta tensiones estructurales entre el uso instrumental de la IA, centrado en eficiencia, adopción y personalización, y el aprendizaje complejo que la educación matemática contemporánea demanda. Identificar, caracterizar y cuantificar esas tensiones a partir de 219 artículos es el aporte central de esta revisión de alcance.

Material y métodos

La investigación adoptó un enfoque cualitativo con diseño exploratorio y descriptivo, materializado en una revisión de alcance (scoping review). Este tipo de diseño permite el mapeo sistemático de un campo de estudio para identificar qué se ha investigado y qué aspectos permanecen sin explorar, sin pretender evaluar la calidad metodológica de los trabajos incluidos. El proceso se desarrolló siguiendo el protocolo PRISMA-ScR (Tricco et al., 2018), actualizado conforme a las orientaciones de la declaración PRISMA 2020 (Page et al., 2021), que orienta cada una de las fases de búsqueda, selección y análisis de la información.

La búsqueda se realizó en febrero de 2026 en seis bases de datos académicas: Web of Science (WoS), Scopus, ERIC, SciELO, IEEE Xplore y ACM Digital Library (ACM DL). Para garantizar una recuperación pertinente y coherente con el objeto de estudio, se diseñaron cadenas booleanas que combinaron tres grupos de descriptores: herramientas de IA, educación matemática y marcos teórico-pedagógicos. El período de cobertura abarcó de enero de 2020 a diciembre de 2025, intervalo que coincide con la consolidación y posterior expansión de las herramientas de IAGen en contextos educativos. La Tabla 1 detalla las cadenas y filtros aplicados en cada base.

Tabla 1
Cadenas de búsqueda y filtros por base de datos

Base	Cadena de búsqueda (campos principales)	Filtros aplicados	n
WoS	TS=((artificial intelligence OR AI) AND (mathematic*) AND (educat* OR pedagog*) AND (framework OR model))	2020 a 2025; Article, Review; English, Spanish; Education y Mathematics	177
Scopus	TITLE-ABS-KEY((artificial intelligence OR Generative AI OR ChatGPT) AND mathematic* AND (theoretical framework OR conceptual framework OR TPACK OR pedagogical))	2020 a 2025; English, Spanish; Computer Science, Social Sciences, Mathematics; Article, Conference paper, Review	133
ERIC	(artificial intelligence OR generative AI OR ChatGPT OR LLM OR intelligent tutoring) AND (mathematics education OR math instruction)	2020 a 2025; Article; English	81
SciELO	((ti:(inteligencia artificial) OR ti:(artificial intelligence)) AND (kw:(matemati* OR mathematic*)) AND ab:(marco teórico OR framework OR teoría OR theory OR modelo))	2020 a 2025; Inglés, Español; Ciencias Sociales, Humanidades	24
IEEE	(Abstract: artificial intelligence OR chatgpt OR large language model) AND (Abstract: mathematic*) AND (Abstract: framework OR model)	2020 a 2025; Conferences, Journals; Mathematics Education	28
ACM DL	Title: (mathematics education OR mathematic*) AND Abstract: (artificial intelligence OR chatgpt OR large language model OR tpack)	2020 a 2025; Research Articles; Journals	14
Total			457

Nota. Las cadenas se adaptaron a la sintaxis particular de cada base de datos. La inclusión de descriptores teórico-pedagógicos (framework, model, TPACK, pedagogical model) en WoS, Scopus y SciELO orientó la recuperación hacia artículos con un lenguaje conceptual explícito. Elaboración propia.

La Tabla 1 evidencia un desequilibrio inherente al diseño: el 67,8 % del corpus inicial proviene de tres bases (WoS, Scopus y ERIC), mientras que SciELO, IEEE y ACM DL aportan en conjunto el 32,2 % restante. Esta asimetría es coherente con la cobertura temática de cada base: WoS, Scopus y ERIC son las bases con mayor volumen de literatura indexada en educación, mientras que las tres restantes tienen perfiles más especializados. Esta composición del corpus inicial debe tenerse en cuenta al interpretar las afirmaciones del estudio, que se refieren al corpus delimitado por los criterios definidos y no al conjunto total de la literatura sobre IA en educación matemática.

Proceso de selección

El proceso siguió las cuatro etapas del flujo PRISMA-ScR. De los 457 registros iniciales se eliminaron 55 duplicados por DOI, título y autor. El cribado de los 402 registros

restantes por título y resumen excluyó 182 estudios bajo tres criterios: IA en disciplinas no educativas (n = 94), tipo de documento no elegible como editoriales, cartas o presentaciones sin revisión por pares (n = 62) e idioma no cubierto (n = 26). Un artículo adicional se excluyó por falta de acceso al texto completo, resultando en un corpus definitivo de 219 artículos. La Tabla 2 sintetiza el flujo por etapas y la Figura 1 lo representa de forma gráfica.

Tabla 2*Flujo de selección PRISMA-ScR del corpus*

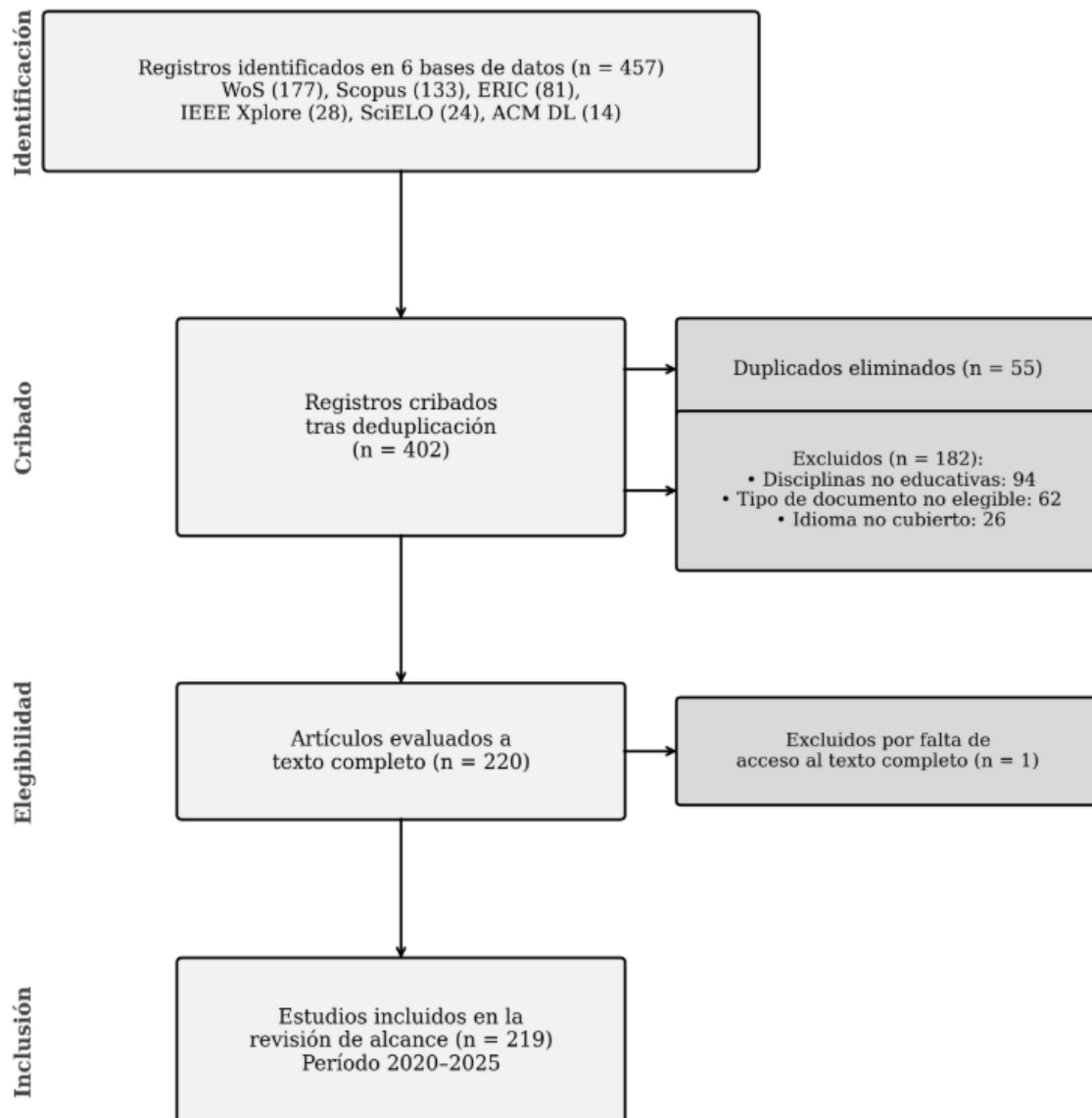
Etapas	Criterio de acción	Artículos	Resultado
Identificación	Búsqueda en 6 bases de datos	457	Corpus inicial
Deduplicación	Eliminación de duplicados por DOI, título y autor	-55	402 registros únicos
Cribado por título y resumen	Exclusión por disciplinas no educativas (n = 94), tipo de documento no elegible (n = 62) e idioma no cubierto (n = 26)	-182	220 artículos elegibles
Texto completo	Exclusión por falta de acceso al texto completo	-1	219 artículos (corpus definitivo)

Nota. *Criterios de inclusión: (a) período 2020 a 2025; (b) relación explícita con IA y educación matemática; (c) artículo empírico, teórico, revisión o estudio de desarrollo con revisión por pares; (d) idioma inglés o español; (e) acceso al texto completo. La inclusión de artículos de revisión dentro del corpus es coherente con el diseño exploratorio. Elaboración propia, adaptado de Tricco et al. (2018) y Page et al. (2021).*

La Tabla 2 muestra que la tasa de retención final del proceso fue del 47,9 % (219 de los 457 registros iniciales), valor coherente con la literatura metodológica para revisiones de alcance en campos emergentes. La exclusión más voluminosa se produjo en el cribado por título y resumen (n = 182), atribuible principalmente a la polisemia del término inteligencia artificial, que en las bases de datos consultadas recupera abundante producción no educativa (disciplinas no educativas concentraron el 51,6 % de las exclusiones de esta etapa).

Figura 1

Diagrama PRISMA-ScR del proceso de selección



Nota. Elaboración propia, adaptado de Tricco et al. (2018) y Page et al. (2021). El proceso fue realizado por el investigador principal; los casos ambiguos de exclusión fueron resueltos mediante consulta al segundo investigador.

Extracción y codificación de los datos

Para cada artículo se construyó una matriz de extracción con 34 variables en seis bloques. Las tres dimensiones cognitivas, modelización, autorregulación y pensamiento crítico, se codificaron mediante un criterio unificado: la dimensión se consideró presente cuando aparecía como proceso matemático declarado en el artículo, como marco teórico central

del estudio (por ejemplo, el Ciclo de Modelización de Borromeo Ferri y colaboradores) o como constructo explícitamente medido o desarrollado; se consideró ausente cuando la mención era incidental o no constituía objeto de la investigación.

La función principal de la IA fue codificada en diez categorías, entre las que destacó la categoría apoyo general al proceso de enseñanza-aprendizaje ($n = 105$; 47,9 %), que agrupa artículos donde la IA actúa como herramienta de soporte sin un rol tecnológico específicamente delimitado. La demanda cognitiva predominante se codificó como alta (razonamiento complejo, modelización, demostración, pensamiento crítico), media (aplicación de procedimientos a contextos nuevos) o baja (reproducción de procedimientos rutinarios), tomando como referencia la taxonomía de tareas matemáticas de Stein y Smith, empleada recientemente para analizar la demanda cognitiva de tareas matemáticas generadas por ChatGPT (Sapkota & Bondurant, 2024)

El proceso fue realizado por el investigador principal con revisión sistemática de los textos completos. Los casos de asignación ambigua, principalmente en demanda cognitiva y función de la IA, fueron consultados con el segundo investigador hasta alcanzar consenso explícito. Esta verificación consensuada afectó aproximadamente al 12 % de las asignaciones de demanda cognitiva y al 8 % de las asignaciones de función de la IA.

Limitaciones del diseño

Se reconocen cuatro limitaciones principales. Primero, sesgo de búsqueda: la inclusión de descriptores teórico-pedagógicos en WoS, Scopus y SciELO privilegia artículos con lenguaje conceptual explícito, por lo que los resultados sobre tipo de estudio, marcos teóricos y herramientas deben interpretarse dentro de este recorte. Segundo, sesgo idiomático: el corpus cubre exclusivamente inglés y español; el portugués fue excluido del diseño de búsqueda, lo que puede subrepresentar producción brasileña o portuguesa no indexada en inglés. Tercero, codificación por evaluador único: la ausencia de un índice de acuerdo intercodificador formal limita la reproducibilidad externa, especialmente en variables interpretativas. Cuarto, heterogeneidad de fuentes: la mezcla de artículos de revista y actas de congresos introduce diferencias en estándares de revisión por pares que no se evaluaron formalmente.

Resultados

Los resultados se presentan en cuatro bloques que responden a las preguntas de investigación: características bibliométricas del corpus, herramientas de IA reportadas, presencia de las tres dimensiones cognitivas complejas y patrones de articulación entre el uso instrumental de la IA y el aprendizaje complejo.

Características bibliométricas del corpus

La distribución anual muestra una aceleración progresiva del campo en el período analizado. Los primeros tres años (2020 a 2022) concentran apenas el 11,4 % de la producción del corpus, mientras que el bienio 2023 a 2024 aporta el 37,0 % y solo el año 2025 absorbe el 51,6 % del total (Tabla 3). Esta concentración en el último año confirma que el campo, en su configuración actual, es un fenómeno fundamentalmente post-2022, asociado a la liberación pública de los modelos conversacionales basados en arquitecturas LLM.

Tabla 3

Distribución anual de los artículos del corpus (N = 219)

Año	n	% del corpus
2020	8	3,7 %
2021	7	3,2 %
2022	10	4,5 %
2023	32	14,6 %
2024	49	22,4 %
2025	113	51,6 %
Total	219	100,0 %

Nota. Elaboración propia a partir de la matriz de extracción.

La Tabla 3 evidencia que el campo se constituyó como tal a partir de 2023, cuando la producción superó por primera vez el 10 % anual. Este patrón temporal coincide con el reportado por Opesemowo y Adewuyi (2024) y Pepin et al. (2025) en revisiones paralelas del mismo período. La concentración del 74,0 % del corpus en el bienio 2024 a 2025 implica que cualquier afirmación sobre el campo se sostiene en literatura de muy corto recorrido, lo cual limita la posibilidad de establecer tendencias consolidadas y subraya la pertinencia de actualizaciones sistemáticas.

En cuanto al tipo de estudio, las revisiones sistemáticas, narrativas o de alcance concentran el 80,4 % del corpus ($n = 176$), seguidas por estudios empíricos cuantitativos (10,5 %), estudios empíricos cualitativos (5,5 %) y propuestas teóricas o de desarrollo (3,6 %). Esta predominancia de revisiones sobre estudios primarios revela un campo que se autointerpreta antes de consolidar evidencia empírica robusta, condición típica de campos emergentes.

Herramientas de IA y funciones reportadas

La Tabla 4 presenta la frecuencia con que aparecen las principales herramientas de IA en el corpus. Las cifras no son mutuamente excluyentes: un mismo artículo puede reportar más de una herramienta.

Tabla 4

Herramientas de IA reportadas en el corpus ($N = 219$)

Herramienta o categoría	n	% del corpus	Tipo principal
ChatGPT / GPT (3.5, 4, 4o)	140	63,9 %	Modelo conversacional generativo
LLM genéricos no especificados	130	59,4 %	Modelos grandes de lenguaje
Sistemas tutores inteligentes (ITS)	74	33,8 %	IA no generativa
Robótica educativa	21	9,6 %	IA no generativa
Otras (Gemini, Claude, Copilot, herramientas propias)	38	17,4 %	Diversas

Nota. Un artículo puede reportar varias herramientas; los porcentajes no suman 100 %. Elaboración propia.

La Tabla 4 confirma la centralidad de ChatGPT como objeto de estudio: el 63,9 % del corpus lo nombra explícitamente, y un 59,4 % adicional refiere a LLM genéricos cuya implementación concreta suele ser ChatGPT u otro sistema funcionalmente equivalente. La presencia simultánea de ITS y robótica educativa (33,8 % y 9,6 % respectivamente) muestra que el campo no se reduce a la IAGen; sin embargo, la lectura cualitativa del corpus evidencia que estas tecnologías no generativas aparecen, en buena medida, como contraste o como contexto histórico del giro hacia los modelos conversacionales. En cuanto a la función reportada de la IA, el 47,9 % de los artículos la describe como apoyo general al proceso de enseñanza-aprendizaje, sin un rol tecnológico específicamente delimitado, mientras que el 89,5 % sitúa al estudiante como usuario activo de la herramienta. Estas cifras, conjuntamente, revelan una alta granularidad descriptiva pero

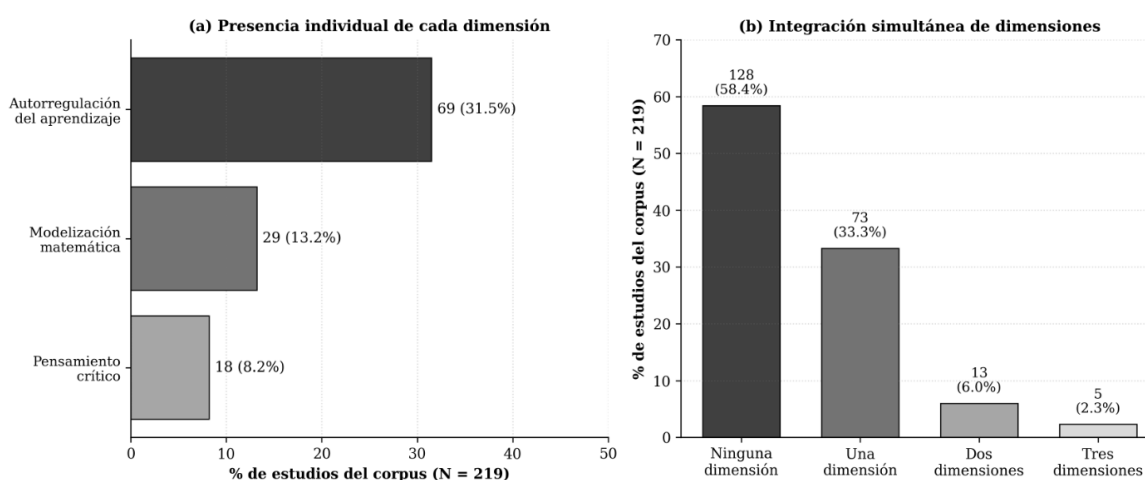
una baja especificidad pedagógica en la caracterización del rol que la IA cumple en los procesos formativos.

Presencia de las tres dimensiones cognitivas complejas

La codificación binaria de las tres dimensiones cognitivas arrojó los resultados sintetizados en la Figura 2.

Figura 2

Presencia de modelización, autorregulación y pensamiento crítico en el corpus



Nota. El panel (a) presenta la presencia individual de cada dimensión; el panel (b) muestra cuántas dimensiones aparecen simultáneamente en cada artículo. Elaboración propia.

El panel (a) de la Figura 2 evidencia un perfil claramente asimétrico. La autorregulación del aprendizaje es la dimensión cognitiva compleja con mayor presencia ($n = 69$; 31,5 %), aunque esta cifra debe matizarse: en aproximadamente dos tercios de los casos se trata de menciones a la autorregulación como condición deseable, no como objeto explícito de diseño didáctico o medición. La modelización matemática aparece en el 13,2 % del corpus ($n = 29$), porcentaje notablemente bajo dado el peso curricular contemporáneo de esta competencia. El pensamiento crítico frente a las respuestas de la IA presenta la cobertura más baja, con apenas el 8,2 % del corpus ($n = 18$).

El panel (b) presenta el hallazgo más significativo del estudio: el 58,4 % del corpus ($n = 128$) no aborda ninguna de las tres dimensiones cognitivas complejas, el 33,3 % ($n = 73$) aborda solo una, el 6,0 % ($n = 13$) integra dos y únicamente el 2,3 % ($n = 5$) integra las

tres simultáneamente. Es decir, en el 91,7 % de los estudios del corpus el aprendizaje cognitivamente complejo aparece como objeto marginal, fragmentario o ausente del diseño de investigación. Este dato configura, a juicio de los autores, la principal brecha del campo en el período analizado.

Procesos matemáticos y demanda cognitiva

La codificación de los procesos matemáticos abordados (Tabla 5) muestra que la resolución de problemas concentra el 77,2 % de los procesos reportados, seguida muy de lejos por el razonamiento y demostración (24,2 %), la representación matemática (16,9 %) y la modelización (13,2 %).

Tabla 5

Procesos matemáticos abordados en el corpus (N = 219)

Proceso matemático	n	% del corpus
Resolución de problemas	169	77,2 %
Razonamiento y demostración	53	24,2 %
Representación matemática	37	16,9 %
Modelización matemática	29	13,2 %
Comunicación y argumentación matemática	22	10,0 %
Conexiones matemáticas	14	6,4 %

Nota. Un artículo puede abordar varios procesos. Elaboración propia.

La Tabla 5 confirma que la resolución de problemas, en su acepción más amplia y muchas veces procedimental, es el proceso matemático dominante en el corpus. La distribución revela una estructura jerárquica donde los procesos de mayor demanda cognitiva (razonamiento, modelización, conexiones) ocupan posiciones progresivamente menores, consolidando la tesis de que el campo se orienta predominantemente al uso instrumental antes que al desarrollo de competencias complejas.

En cuanto a la demanda cognitiva predominante, el 47,5 % de los estudios (n = 104) presenta una demanda media (aplicación de procedimientos a contextos nuevos), el 32,4 % (n = 71) una demanda baja (reproducción procedimental) y solo el 20,1 % (n = 44) una demanda alta (razonamiento complejo, modelización, demostración o pensamiento crítico). La conjunción de la estructura procesal y la demanda cognitiva refuerza la lectura central: el corpus, en su conjunto, opera en una zona cognitiva media-baja.

Tensiones estructurales entre uso instrumental y aprendizaje complejo

A partir del análisis cualitativo de los datos cuantitativos anteriores, se identifican cinco tensiones estructurales que caracterizan al campo en el período analizado. La Tabla 6 las sintetiza.

Tabla 6

Cinco tensiones estructurales identificadas en el corpus

#	Tensión	Indicador cuantitativo en el corpus
1	Eficiencia instruccional frente a lucha cognitiva productiva	IA como apoyo general (47,9 %) frente a aprendizaje complejo (2,3 % integra las tres dimensiones)
2	Resolución procedimental frente a modelización matemática	Resolución de problemas (77,2 %) frente a modelización (13,2 %)
3	Personalización adaptativa frente a autorregulación del aprendizaje	Autorregulación presente solo en el 31,5 %; aparece como condición declarativa antes que como objeto de diseño
4	Retroalimentación inmediata frente a pensamiento crítico	Pensamiento crítico presente solo en el 8,2 % del corpus
5	Adopción tecnológica frente a fundamentación didáctica	Predominio de revisiones (80,4 %) y baja granularidad pedagógica en el 47,9 % codificado como apoyo general

Nota. Elaboración propia a partir del análisis del corpus.

La Tabla 6 articula los hallazgos cuantitativos en una lectura estructural del campo: cada tensión documenta una asimetría sistemática entre lo que la literatura prioriza y lo que las competencias matemáticas contemporáneas demandan. La presencia simultánea de estas cinco tensiones, sostenida por evidencia cuantitativa convergente, constituye el principal aporte analítico de la presente revisión.

Discusión

Principios y generalizaciones derivados del corpus

Tres principios emergen como generalizaciones sostenibles del análisis. Primero, la literatura sobre IA Gen en educación matemática del período 2020 a 2025 está dominada por una lógica de adopción instrumental que reproduce, ampliándolas, las orientaciones de eficiencia previas a la irrupción de los modelos conversacionales. La concentración del 47,9 % de los estudios en la categoría apoyo general al proceso de enseñanza-aprendizaje no es un accidente metodológico sino una expresión cualitativa del modo en que el campo conceptualiza el aporte de la IA: como soporte difuso a la productividad instruccional antes que como mediación cognitiva específica.

Segundo, el aprendizaje cognitivamente complejo, entendido como la integración explícita de modelización, autorregulación y pensamiento crítico, constituye una zona marginal del campo: apenas el 2,3 % del corpus aborda las tres dimensiones de manera simultánea, y el 58,4 % no aborda ninguna. Este principio establece una asimetría que opera en dos niveles: a nivel cuantitativo, las dimensiones complejas son objeto residual; a nivel cualitativo, cuando aparecen lo hacen mayoritariamente como condiciones declarativas, no como objeto explícito de diseño didáctico, lo cual reduce su contribución formativa real.

Tercero, la jerarquía de procesos matemáticos del corpus consagra a la resolución de problemas (77,2 %) como proceso hegemónico, pero esta hegemonía conviene leerse con cautela: en buena parte del corpus la resolución de problemas se operacionaliza en su acepción procedimental, próxima a la aplicación de algoritmos, antes que en su acepción didáctica plena, que demanda heurística, validación y conexión con contextos auténticos. La distancia entre el 77,2 % de resolución de problemas y el 13,2 % de modelización ilustra este punto: cuando la IA participa, lo hace facilitando la ejecución antes que el proceso completo de problematizar, modelar y validar.

Concordancias con trabajos previos

Los hallazgos del presente estudio dialogan con tres revisiones recientes del campo. Nguyen y Pham (2025), en su revisión de 58 artículos para el período 2015 a 2025,

reportaron una concentración en tutoría adaptativa e ITS con escasa atención a procesos de alta demanda cognitiva. Los datos del presente corpus replican y refinan ese hallazgo: la demanda cognitiva alta aparece solo en el 20,1 % de los estudios y la integración de las tres dimensiones complejas en el 2,3 %, lo que cuantifica con mayor precisión la brecha que Nguyen y Pham documentaron de forma cualitativa.

Panqueban y Huincahue (2024) constataron, sobre 29 artículos, la hegemonía de marcos tecnológicos sobre marcos propios de la didáctica matemática. El corpus de la presente revisión amplía esa observación: la presencia de revisiones sistemáticas (80,4 %) y la baja granularidad pedagógica del 47,9 % codificado como apoyo general son indicadores cuantitativos coherentes con la hegemonía tecnológica que Panqueban y Huincahue identificaron. La concordancia entre ambos estudios, basados en corpus de tamaños muy diferentes (29 frente a 219), sugiere que el patrón observado no obedece únicamente al sesgo de búsqueda del presente estudio sino a una estructura compartida del campo.

Yi et al. (2025), en su metaanálisis centrado en la efectividad de la IA sobre el rendimiento matemático en K-12, privilegiaron variables de desempeño académico antes que el desarrollo de competencias complejas. Esta orientación es coherente con la tensión 1 identificada en el presente estudio (eficiencia instruccional frente a lucha cognitiva productiva): el campo, en su mayoría, opera bajo un supuesto implícito según el cual la efectividad de la IA se mide por su capacidad para mejorar el desempeño en tareas existentes, antes que por su capacidad para enriquecer el repertorio cognitivo del estudiante. Pepin et al. (2025) y Opesemowo y Adewuyi (2024) llegaron a conclusiones convergentes en revisiones paralelas, lo que refuerza la consistencia del patrón.

Los estudios producidos en el contexto costarricense, tanto la experiencia de Acuña y Marrero (2025) en un curso de Cálculo como las propuestas de innovación didáctica de Ramírez-Bogantes et al. (2026) y el estudio de actitudes estudiantiles de Meza-Cascante et al. (2026), confluyen en una observación cualitativa: la integración de IAGen en educación matemática universitaria es viable, pero exige una mediación pedagógica explícita que el corpus internacional, según se ha mostrado, todavía no ha consolidado.

Consecuencias teóricas y aplicaciones prácticas

Desde el plano teórico, los hallazgos sugieren que el campo de la IA en educación matemática necesita ampliar su anclaje didáctico-disciplinar. La hegemonía de marcos genéricos de integración tecnológica, tipo TPACK o sus extensiones, debe complementarse con marcos propios de la didáctica de la matemática que tematizen específicamente cómo el razonamiento matemático, la modelización y la validación lógica se reconfiguran cuando un agente artificial puede producir soluciones plausibles pero potencialmente erróneas. El enfoque instrumental de Rabardel, la idoneidad didáctica del enfoque ontosemiótico (Pochulu y Font, 2025) y los desarrollos recientes sobre la actividad matemática mediada por IAGen (Yoon et al., 2024) ofrecen vías productivas que la literatura analizada solo incipientemente ha incorporado.

Desde el plano práctico, las cinco tensiones identificadas configuran una agenda concreta para el diseño didáctico. Primero, ante la tensión entre eficiencia y lucha cognitiva productiva, los docentes deben diseñar tareas que aprovechen deliberadamente las limitaciones de la IA, sus errores y sus alucinaciones, para promover la validación crítica por parte del estudiante (Urhan et al., 2024; Yoon et al., 2024). Segundo, frente a la baja presencia de modelización (13,2 %), los programas formativos deben incorporar de forma sistemática experiencias de modelización donde la IA opere como apoyo en una fase del ciclo modelizador (Greefrath y Vorhölter, 2022), no como sustituto del proceso completo.

Tercero, frente al carácter declarativo predominante de la autorregulación del aprendizaje, las propuestas didácticas deben operacionalizar esta dimensión mediante guiones de interacción explícitos con la IA, donde el estudiante planifique, monitoree y evalúe su uso de la herramienta (Contel y Cusi, 2025; Panadero y Lipnevich, 2022). Cuarto, frente a la baja presencia de pensamiento crítico (8,2 %), conviene incorporar a la evaluación académica criterios explícitos sobre la capacidad del estudiante para detectar errores, validar argumentos y auditar las soluciones que la IA produce. Quinto, frente a la tensión entre adopción tecnológica y fundamentación didáctica, la formación docente debe priorizar el desarrollo de criterios didácticos para decidir cuándo y cómo integrar la IAGen, antes que la familiarización con herramientas específicas que el mercado actualiza con frecuencia trimestral.

Vacíos de investigación detectados

Más allá de las tensiones, el análisis del corpus permite identificar cuatro vacíos de investigación que configuran una agenda prioritaria. El primero es la ausencia de investigación empírica longitudinal sobre los efectos del uso sostenido de IAGen en el desarrollo de competencias matemáticas complejas: el 80,4 % del corpus son revisiones y solo el 16,0 % son estudios empíricos primarios. El segundo es la subrepresentación de la producción iberoamericana indexada en bases internacionales, lo que dificulta transferir hallazgos a contextos curriculares y culturales específicos. El tercero es la escasez de estudios que articulen explícitamente las tres dimensiones cognitivas complejas, vacío que limita la posibilidad de elaborar propuestas didácticas integradas. El cuarto es la ausencia de marcos teóricos propios para conceptualizar la actividad matemática del estudiante cuando un agente artificial puede producir soluciones plausibles y, por tanto, reconfigura las condiciones epistemológicas del aprendizaje.

Conclusiones

La revisión de alcance de 219 artículos publicados entre 2020 y 2025 sobre IA en educación matemática permite formular tres conclusiones principales y un conjunto de recomendaciones para la investigación futura y la práctica docente.

Primera conclusión: el campo se ha expandido de manera exponencial pero asimétrica. La concentración del 51,6 % del corpus en 2025 y del 74,0 % en el bienio 2024 a 2025, junto con la hegemonía de ChatGPT (63,9 %) y de los LLM genéricos (59,4 %), configura un campo emergente cuya consolidación teórica es aún incipiente. El predominio de revisiones sistemáticas (80,4 %) sobre estudios empíricos primarios evidencia una literatura que se autointerpreta antes de haber acumulado evidencia robusta, condición que demanda actualización periódica de los mapeos realizados.

Segunda conclusión: las tres dimensiones cognitivas complejas que la educación matemática contemporánea privilegia, modelización, autorregulación y pensamiento crítico, ocupan una zona marginal del campo. La autorregulación aparece en el 31,5 % del corpus pero predominantemente como condición declarativa; la modelización en el 13,2 % y el pensamiento crítico en apenas el 8,2 %. Solo el 2,3 % del corpus integra las

tres dimensiones de manera simultánea, y el 58,4 % no aborda ninguna. Esta distribución configura la principal brecha del campo y revela que, en el período analizado, la investigación se ha alejado de las competencias formativas más exigentes para concentrarse en usos instrumentales de la IA.

Tercera conclusión: el corpus revela cinco tensiones estructurales entre el uso instrumental de la IA y el aprendizaje complejo: eficiencia instruccional frente a lucha cognitiva productiva, resolución procedimental frente a modelización, personalización adaptativa frente a autorregulación, retroalimentación inmediata frente a pensamiento crítico, y adopción tecnológica frente a fundamentación didáctica. Estas tensiones no son meramente conceptuales sino que aparecen sostenidas por evidencia cuantitativa convergente del corpus y configuran, en su conjunto, la frontera analítica del campo.

Recomendaciones

De las conclusiones anteriores se derivan cuatro recomendaciones principales. Para la investigación: priorizar estudios empíricos longitudinales que evalúen el efecto del uso sostenido de IAGen sobre el desarrollo de competencias matemáticas complejas, particularmente en contextos iberoamericanos donde la producción es subrepresentada. Para la formación docente: incorporar sistemáticamente criterios didácticos explícitos para decidir cuándo y cómo integrar la IAGen, antes que la familiarización con herramientas específicas, e incluir formación sobre los sesgos sistemáticos y las alucinaciones que los LLM producen al abordar contenidos matemáticos. Para el diseño didáctico: diseñar tareas que aprovechen deliberadamente las limitaciones de la IA como recurso para activar la validación crítica del estudiante, e incorporar la modelización, la autorregulación y el pensamiento crítico como objetos explícitos de diseño, no como condiciones declarativas. Para la evaluación: ampliar los criterios evaluativos para incluir la capacidad del estudiante de auditar, validar o refutar las soluciones producidas por la IA, en línea con la reconfiguración del razonamiento matemático que la presencia de agentes artificiales productores de soluciones plausibles introduce.

Las limitaciones del estudio, sesgo de búsqueda hacia descriptores teórico-pedagógicos, exclusión del portugués, codificación por evaluador único y heterogeneidad de fuentes, demarcan el alcance de las afirmaciones. Aun así, la consistencia de los hallazgos con revisiones independientes (Nguyen y Pham, 2025; Opesemowo y Adewuyi, 2024;



Panqueban y Huincahue, 2024; Pepin et al., 2025) sugiere que el patrón identificado, predominio del uso instrumental sobre el aprendizaje complejo, caracteriza al campo en su conjunto y no únicamente al corpus aquí analizado.

Referencias bibliográficas

- Acuña, C. R. Y., y Marrero, C. M. (2025). Integración de metodologías activas e inteligencia artificial en la enseñanza del tema de sucesiones y series: una experiencia de clase en un curso semipresencial del Instituto Tecnológico de Costa Rica. Cuadernos de Investigación y Formación en Educación Matemática, 18(2), 135-160. <https://doi.org/10.15517/eqxwfy21>
- Borromeo Ferri, R. (2021). Learning how to teach mathematical modeling in school and teacher education. Springer. <https://doi.org/10.1007/978-3-319-68072-9>
- Campos, M. J. E. (2024). Inteligencia artificial en la educación superior de Costa Rica: desafíos y oportunidades desde una perspectiva ética [Tesis de Maestría, Universidad Nacional de Costa Rica]. Repositorio Institucional UNA. <https://hdl.handle.net/11056/29614>
- Chacón-Rivadeneira, K., Morales-Maure, L., García-Marimón, O., Sáez-Delgado, F., Gutiérrez González, J., y Alfaro Ponce, B. (2024). Artificial intelligence adoption in Latin American mathematics education: challenges and opportunities. Journal of Posthumanism, 4(3), 1141-1161. <https://doi.org/10.63332/joph.v4i3.3195>
- Contel, F., & Cusi, A. (2025). *Investigating the role of ChatGPT in supporting metacognitive processes during problem-solving activities*. Digital Experiences in Mathematics Education, 11, 167-191. <https://doi.org/10.1007/s40751-024-00164-7>
- Greefrath, G., y Vorhölter, K. (2022). Teaching and learning mathematical modelling: approaches and developments from recent research. ZDM Mathematics Education, 54, 1-15. <https://doi.org/10.1007/s11858-021-01330-6>
- Meza-Cascante, L. G., Ramírez-Bogantes, M., y Meza-Chavarría, L. A. (2026). Actitudes del estudiantado de ingeniería ante la inteligencia artificial generativa: estudio en el Instituto Tecnológico de Costa Rica. Revista Tecnología en Marcha, 39(5), 137-146. <https://doi.org/10.18845/tm.v39i5.8502>
- Nguyen, D. T., y Pham, Q. V. (2025). The evolving landscape of AI integration in mathematics education: a systematic review of trends (2015-2025). Eurasia Journal of Mathematics,



- Science and Technology Education, 21(10), em2714.
<https://doi.org/10.29333/ejmste/17078>
- Opesemowo, O. A. G., y Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: the emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(7), em2478. <https://doi.org/10.29333/ejmste/14762>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Panadero, E., y Lipnevich, A. A. (2022). A review of feedback models and typologies: towards an integrative model of feedback elements. *Educational Research Review*, 35, 100416. <https://doi.org/10.1016/j.edurev.2021.100416>
- Panqueban, J., y Huincahue, J. (2024). Artificial intelligence in mathematics education: a systematic review. *Uniciencia*, 38(1). <https://doi.org/10.15359/ru.38-1.20>
- Pepin, B., Buchholtz, N., y Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 11(1), 9-41. <https://doi.org/10.1007/s40751-025-00172-1>
- Pochulu, M. D., y Font, V. (2025). Idoneidad didáctica de tareas de matemáticas reformuladas con inteligencia artificial. *Paradigma*, 46(1), e2025027. <https://doi.org/10.37618/PARADIGMA.1011-2251.2025.e2025027.id1621>
- Ramírez-Bogantes, M., Monge-Fallas, J., Borbón-Alpízar, A., Gutiérrez-Montenegro, M. V., Acuña-Chacón, R., y Meza-Cascante, L. G. (2026). Innovación didáctica con IA generativa en matemática universitaria: experiencias en el TEC. *Revista Tecnología en Marcha*, 39(5), 147-156. <https://doi.org/10.18845/tm.v39i5.8503>
- Sapkota, B., & Bondurant, L. (2024). Assessing concepts, procedures, and cognitive demand of ChatGPT-generated mathematical tasks. *International Journal of Technology in Education*, 7(2), 218-238. <https://doi.org/10.46328/ijte.677>
- 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>

Urhan, S., Gençaslan, O., y Dost, Ş. (2024). An argumentation experience regarding concepts of calculus with ChatGPT. *Interactive Learning Environments*, 1–26. <https://doi.org/10.1080/10494820.2024.2308093>

Yi, L., Liu, D., Jiang, T., & Xian, Y. (2025). The effectiveness of AI on K-12 students' mathematics learning: A systematic review and meta-analysis. *International Journal of Science and Mathematics Education*, 23(4), 1105–1126. <https://doi.org/10.1007/s10763-024-10499-7>

Yoon, H., Hwang, J., Lee, K., Roh, K. H., y Kwon, O. N. (2024). Students' use of generative artificial intelligence for proving mathematical statements. *ZDM Mathematics Education*, 56(7), 1531–1551. <https://doi.org/10.1007/s11858-024-01629-0>

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Agradecimientos:

No aplica.

Nota sobre el uso de IA en la elaboración del manuscrito:

Durante la elaboración del manuscrito, los autores utilizaron asistentes basados en inteligencia artificial generativa para apoyar tareas de revisión de redacción, organización de tablas y verificación cruzada de cifras frente a la matriz de extracción.

La interpretación, el análisis y las conclusiones son responsabilidad exclusiva de los autores.

Nota:

El artículo no es producto de una publicación anterior.