

<https://doi.org/10.69639/arandu.v13i2.2197>

Entre el algoritmo y la comprensión: una revisión exploratoria de la inteligencia artificial en la enseñanza de la matemática (2020–2025)

Between the algorithm and understanding: a scoping review of artificial intelligence in mathematics education (2020–2025)

Kenneth Mejía Luna

kmluna16@estudiantec.cr

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

Licenciatura en Enseñanza de la Matemática
Tecnológico de Costa Rica
Cartago – Costa Rica

Melvin Ramírez Bogantes

meramirez@itcr.ac.cr

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

Profesor investigador | Escuela de Matemática
Tecnológico de Costa Rica
Heredia – Costa Rica

Artículo recibido: 18 marzo 2026- Aceptado para publicación: 20 abril 2026
Conflictos de intereses: Ninguno que declarar.

RESUMEN

El presente artículo sistematiza y clasifica los marcos teóricos sobre inteligencia artificial (IA) en educación publicados entre 2020 y 2025, con énfasis en su aplicabilidad a la enseñanza de la matemática. Se condujo una revisión exploratoria (Scoping Review) conforme a la declaración PRISMA-ScR, que recuperó 457 registros en Web of Science, Scopus, ERIC, IEEE Xplore, SciELO y ACM Digital Library. Tras el cribado en dos etapas se incluyeron 153 estudios, de los cuales 112 conforman el subcorpus de enseñanza de la matemática y 80 constituyen el subcorpus efectivo con análisis didáctico operativo. Los estudios se analizaron mediante una matriz de extracción de 73 variables organizada en cuatro núcleos conceptuales (tecnológico, pedagógico, crítico-ético y psicológico-afectivo) cruzados con tres dimensiones transversales (función del marco, nivel de generalidad y relación con la matemática). Los resultados evidencian un predominio del núcleo tecnológico (49,0 %) seguido del pedagógico (30,1 %), con la articulación entre ambos concentrando el 48,7 % de las 158 co-ocurrencias identificadas. TPACK y sus extensiones constituyen el marco más recurrente (14,4 %), mientras que los marcos críticos y éticos se consolidan como núcleo emergente (13,7 %) aunque con predominio analítico (81,0 %) y escasa proyección hacia el diseño. En el subcorpus efectivo, el razonamiento y demostración (70,0 %) y el pensamiento crítico orientado a validar las respuestas de la IA (67,5 %) emergen como los procesos matemáticos más promovidos, mientras que la modelización matemática (21,2

%) permanece subrepresentada. Se documenta así un desplazamiento pedagógico relevante: el razonamiento matemático con IA se ejerce, en buena medida, como razonamiento crítico sobre la IA. Se proponen criterios operativos para el diseño didáctico, la mediación docente y la evaluación del aprendizaje matemático mediado por IA.

Palabras clave: inteligencia artificial en educación, revisión exploratoria, marcos teóricos, enseñanza de la matemática, TPACK

ABSTRACT

This article systematizes and classifies the theoretical frameworks on artificial intelligence (AI) in education published between 2020 and 2025, with an emphasis on their applicability to mathematics teaching. A Scoping Review was conducted following the PRISMA-ScR statement, which retrieved 457 records from Web of Science, Scopus, ERIC, IEEE Xplore, SciELO, and ACM Digital Library. After a two-stage screening, 153 studies were included; 112 constitute the mathematics subcorpus and 80 the effective subcorpus with operational didactic analysis. Studies were analyzed through a 73-variable extraction matrix organized around four conceptual nuclei (technological, pedagogical, critical-ethical, and psychological-affective) crossed with three transversal dimensions (framework function, level of generality, and relationship to mathematics). The results reveal a predominance of the technological nucleus (49.0 %), followed by the pedagogical one (30.1 %), with the articulation between both concentrating 48.7 % of the 158 identified co-occurrences. TPACK and its extensions constitute the most recurrent framework (14.4 %), while critical-ethical frameworks consolidate as an emerging nucleus (13.7 %) though with analytical predominance (81.0 %) and limited projection toward design. In the effective subcorpus, reasoning and proof (70.0 %) and critical thinking aimed at validating AI responses (67.5 %) emerge as the most promoted mathematical processes, while mathematical modeling (21.2 %) remains underrepresented. A relevant pedagogical shift is thereby documented: mathematical reasoning with AI is exercised, to a significant extent, as critical reasoning about AI. Operational criteria are proposed for instructional design, teaching mediation, and assessment of mathematical learning mediated by AI.

Keywords: artificial intelligence in education, scoping review, theoretical frameworks, mathematics teaching, TPACK

Todo el contenido de la Revista Científica Internacional Arandu UTIC publicado en este sitio está disponible bajo licencia Creative Commons Attribution 4.0 International. 

INTRODUCCIÓN

La inteligencia artificial (IA) ha dejado de ser un recurso periférico para convertirse en un entorno habitual de estudio, trabajo docente y toma de decisiones institucionales (Pepin et al., 2025). En el campo específico de la enseñanza de la matemática, esta transformación introduce retos distintivos: garantizar el sentido formativo de las tareas, preservar la calidad del razonamiento y sostener la validez de la evaluación cuando intervienen sistemas automatizados capaces de producir soluciones plausibles pero lógicamente inválidas (Wardat et al., 2023; Yoon et al., 2024). Frente a este escenario, la descripción de herramientas no basta: se requiere una fundamentación teórica rigurosa que permita comprender por qué y cómo integrar la IA en relación con objetivos, contenidos y mediación docente.

La producción académica del periodo 2020–2025 ha crecido con una rapidez inusual, especialmente tras la irrupción de los modelos de lenguaje generativos a finales de 2022 (Opesemowo y Adewuyi, 2024). Sin embargo, este crecimiento ha ocurrido de manera fragmentada: coexisten marcos de distinta naturaleza —pedagógicos, psicológicos y afectivos, tecnológicos, y críticos o éticos— sin una sistematización rigurosa que los organice, compare y evalúe. Esta dispersión dificulta construir una base conceptual integrada que oriente tanto la investigación como la toma de decisiones didácticas, y es particularmente problemática en la educación matemática, donde la literatura que articula de manera explícita estos marcos con el aprendizaje disciplinar es aún escasa (Ning et al., 2024; Priyanda et al., 2025).

En el contexto costarricense, la pertinencia del problema se acentúa. Campos (2024) documentó que el sistema educativo nacional requiere con urgencia directrices pedagógicas y éticas claras que regulen la adopción tecnológica, particularmente tras las brechas agravadas por la pandemia. A nivel institucional, Acuña y Marrero (2025) evidenciaron 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 de las herramientas. En esta misma línea, Ramírez-Bogantes et al. (2026) destacan que la innovación didáctica con inteligencia artificial generativa en la matemática universitaria exige ir más allá de la adopción técnica, posicionando a la tecnología como una mediadora cognitiva y ética. Asimismo, para que esta integración teórica y pedagógica logre consolidarse, resulta indispensable comprender las actitudes que el propio estudiantado manifiesta hacia estas tecnologías emergentes en el aula (Meza-Cascante et al., 2026).

A partir de este diagnóstico, la investigación que da origen al presente artículo se propuso como objetivo general realizar una revisión exploratoria siguiendo la declaración PRISMA-ScR de la literatura publicada entre 2020 y 2025, con el fin de mapear y sistematizar los marcos teóricos sobre IA aplicados a la docencia y valorar su aplicabilidad para la enseñanza de la matemática. La pregunta central que orientó el estudio fue: ¿cómo han sido utilizados, clasificados

y problematizados los marcos teóricos sobre IA en la docencia durante 2020–2025, y cuál es su pertinencia y aplicabilidad para la enseñanza de la matemática? Este artículo presenta los principales resultados de esa revisión, con particular énfasis en los hallazgos del subcorpus de enseñanza de la matemática.

MATERIALES Y MÉTODOS

El estudio adoptó un enfoque de revisión exploratoria (Scoping Review) de carácter cualitativo-descriptivo, siguiendo los lineamientos de la declaración PRISMA-ScR (Tricco et al., 2018) y la actualización PRISMA 2020 (Page et al., 2021). A diferencia de las revisiones sistemáticas orientadas a estimar efectos, este diseño permite mapear conceptos clave, identificar vacíos y caracterizar la evidencia en un campo emergente (Arksey y O'Malley, 2005). La ventana temporal comprendió publicaciones entre el 1 de enero de 2020 y el 31 de diciembre de 2025.

Fuentes de información y estrategia de búsqueda

La búsqueda se ejecutó en seis bases de datos seleccionadas por su cobertura internacional y pertinencia para el campo: Web of Science Core Collection, Scopus, ERIC (vía ProQuest), IEEE Xplore, SciELO y ACM Digital Library. Adicionalmente, se realizó rastreo de referencias hacia atrás y hacia adelante mediante Google Scholar. La estrategia combinó sistemáticamente términos controlados y palabras clave en tres bloques temáticos: (1) inteligencia artificial, (2) educación y enseñanza, y (3) marcos teóricos, empleando operadores booleanos y truncadores adaptados a la sintaxis de cada base.

Criterios de elegibilidad

Se incluyeron los estudios que (a) declararon explícitamente el uso de una teoría o modelo sobre IA en educación, (b) operacionalizaron dicha teoría en el diseño instruccional, en los instrumentos o en las categorías de análisis (Creswell y Creswell, 2018), (c) se centraron en procesos de enseñanza, aprendizaje o gestión educativa en niveles formales y (d) fueron publicados como artículos revisados por pares, actas de conferencias o capítulos de libro, en inglés o español. Se excluyeron los trabajos técnicos enfocados exclusivamente en el desarrollo algorítmico sin análisis pedagógico, las publicaciones de opinión sin metodología sistemática y los estudios donde el marco teórico aparecía de manera puramente nominal sin guiar el diseño ni el análisis.

Proceso de selección y extracción

El cribado se ejecutó en dos etapas secuenciales en Rayyan: un cribado preliminar de títulos y resúmenes y un cribado a texto completo. Para calibrar la aplicación operativa de los criterios, se realizó un pilotaje inicial con una muestra aleatoria de 50 registros. La extracción de datos se sistematizó en una matriz estructurada de 73 variables organizadas en siete bloques (identificación bibliográfica, contexto, marco teórico, operacionalización, aplicación didáctica, criterios éticos y valoración de pertinencia). La síntesis se ejecutó en tres niveles: un mapeo

general sobre la totalidad del corpus ($n = 153$), un análisis del subcorpus específico de enseñanza de la matemática ($n = 112$) y una profundización sobre el subcorpus efectivo con análisis didáctico operativo ($n = 80$), al que se aplicó análisis de contenido cualitativo para identificar mediaciones didácticas, procesos matemáticos promovidos y criterios de evaluación reportados.

Esquema analítico

Los marcos identificados se clasificaron en cuatro núcleos conceptuales interrelacionados: tecnológico (modelos de aceptación e integración como TAM, UTAUT, TPACK y AI-TPACK); pedagógico (enfoques sobre el aprendizaje y la mediación didáctica, incluidos marcos específicos de la didáctica matemática como el enfoque instrumental, APOS, RME e Idoneidad Didáctica del EOS); crítico y ético (sesgo algorítmico, explicabilidad, privacidad, integridad académica y equidad); y psicológico-afectivo (motivación, autoeficacia, autorregulación del aprendizaje y carga cognitiva). Cada marco se categorizó adicionalmente según tres dimensiones transversales: (1) función en el estudio —fundamentación teórica, diseño didáctico, análisis e interpretación o síntesis—; (2) nivel de generalidad —macro (metateorías), meso (modelos de integración) o micro (constructos específicos)—; y (3) relación con la enseñanza de la matemática —específica de la disciplina, transferida desde otro dominio, o no aplicable—.

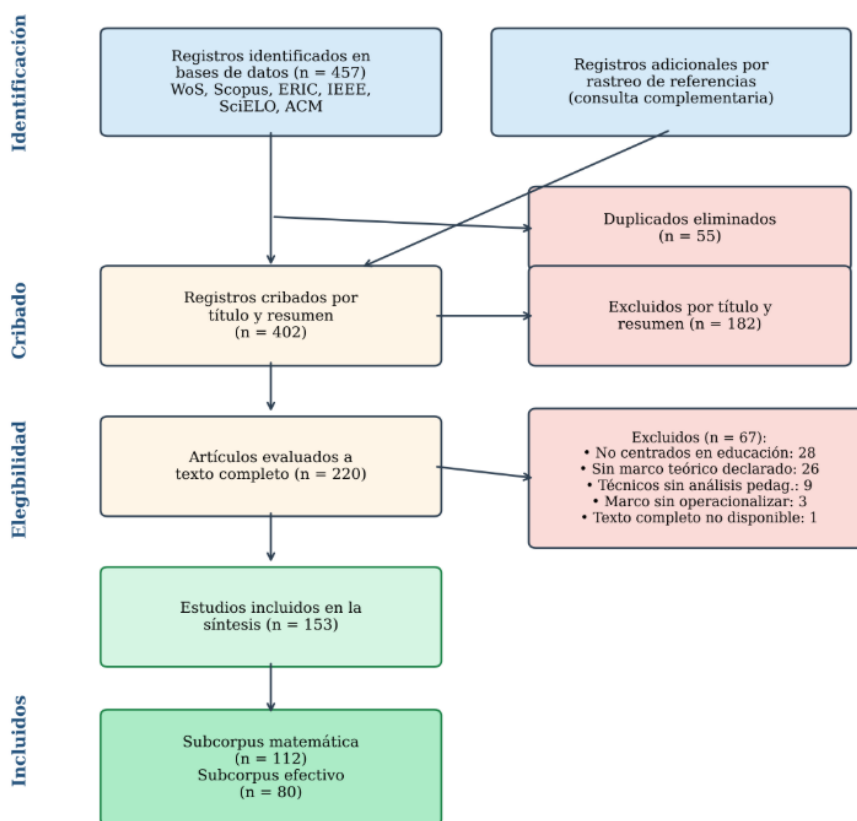
RESULTADOS Y DISCUSIÓN

Flujo de selección y composición del corpus

La búsqueda sistemática recuperó 457 registros en las seis bases de datos consultadas. Tras la eliminación de 55 duplicados, se cribaron 402 registros por título y resumen y se evaluaron 220 a texto completo. La aplicación estricta del criterio de operacionalización del marco teórico condujo a la exclusión de 67 estudios en esta última etapa, configurando un corpus definitivo de 153 estudios (tasa de retención global del 33,5 %). De ellos, 112 (73,2 %) conforman el subcorpus de enseñanza de la matemática y 80 (71,4 % del subcorpus matemática) constituyen el subcorpus efectivo con análisis didáctico operativo. La Figura 1 sintetiza el flujo completo.

Figura 1

Diagrama de flujo PRISMA-ScR del proceso de selección de estudios



Nota. Elaboración propia a partir de la matriz de extracción, adaptado de Tricco et al. (2018) y Page et al. (2021).

Un hallazgo significativo del proceso de selección es que el motivo más frecuente de exclusión en la etapa de texto completo ($n = 28$; 41,8 %) no fue la ausencia de marco teórico sino la falta de centramiento en procesos de enseñanza y aprendizaje. Esto documenta un fenómeno específico del campo: una fracción significativa de la literatura etiquetada como «IA en educación» aborda en realidad contextos para-educativos o tangenciales. El agregado de las exclusiones relativas al uso nominal o técnico del marco (motivos 2, 3 y 4) suma 38 exclusiones (56,7 % del total), lo que ofrece sustento empírico para la hipótesis que orientó el estudio: en el periodo 2020–2025 una fracción considerable de la literatura sobre IA en educación adolece de fundamentación teórica explícita o la instrumentaliza superficialmente.

Tabla 1

Rendimiento por base de datos a lo largo del flujo PRISMA-ScR

Base de datos	Identif.	Dedup.	Cribado T/R	Texto compl.	Definit.	Tasa (%)
Web of Science (WoS)	177	-7	-61	109	83	46,9
Scopus	133	-14	-69	50	36	27,1
ERIC (vía ProQuest)	81	-27	-22	32	23	28,4
IEEE Xplore	28	-7	-9	12	10	35,7

SciELO	24	0	-10	14	0	0,0
ACM Digital Library	14	0	-11	3	1	7,1
Total (combinado)	457	-55	-182	220	153	33,5

Nota. La tasa de retención final se calcula como (Definitivo / Identificados) × 100. Elaboración propia a partir de la matriz de extracción.

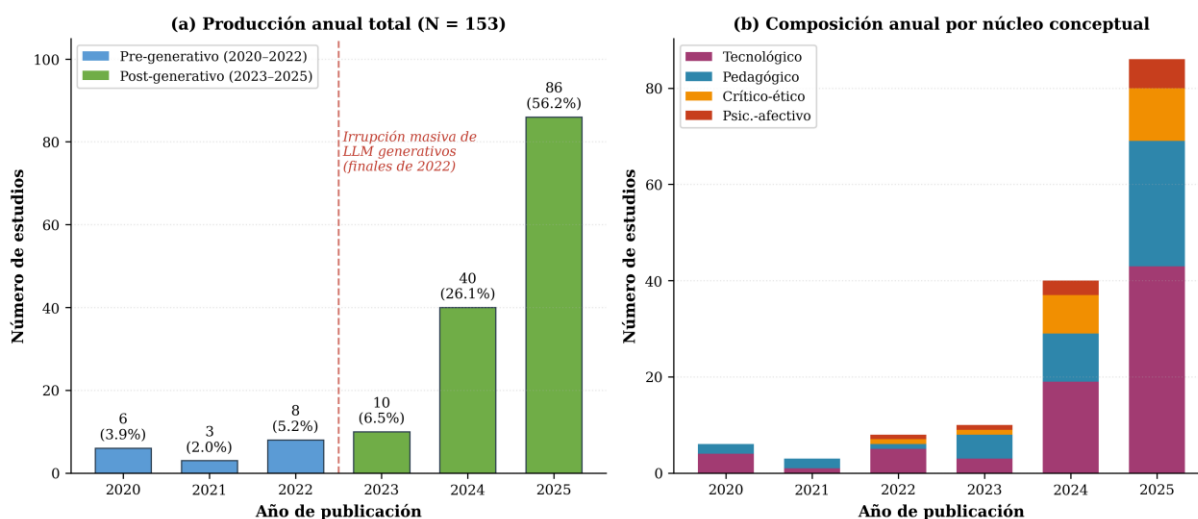
Distribución temporal y dinámica del campo

La distribución anual del corpus (Figura 2, panel a) muestra una asimetría muy marcada. El 82,4 % del corpus se concentra en 2024–2025 (126 de 153 estudios), mientras que el trienio 2020–2022 aporta solo el 11,1 % (17 estudios). Este patrón es consistente con el punto de inflexión que marcó la democratización de los modelos de lenguaje generativos a finales de 2022: la presencia masiva de ChatGPT y herramientas análogas reconfiguró tanto la investigación académica como las prácticas de aula (Opesemowo y Adewuyi, 2024; Pepin et al., 2025). Solo en 2025 se publicaron 86 estudios, equivalentes al 56,2 % del corpus total.

El panel (b) de la Figura 2 descompone esta producción por núcleo conceptual y revela una segunda asimetría: el núcleo tecnológico no solo es el más frecuente en el agregado sino también el primer responsable del crecimiento. En 2025 concentra 43 de los 86 estudios del año (50,0 %), seguido del pedagógico con 26 (30,2 %). El núcleo crítico-ético, casi ausente entre 2020 y 2022, crece sostenidamente a partir de 2024 y alcanza 11 estudios en 2025 (12,8 % del año), lo que refleja la maduración de las preocupaciones éticas del campo en paralelo a la expansión de la IA generativa.

Figura 2

Distribución temporal del corpus y composición por núcleo conceptual (2020–2025)



Nota. N = 153. Los porcentajes del panel (a) se calculan sobre el total del corpus. El panel (b) descompone el total anual según el núcleo predominante de cada estudio. Elaboración propia a partir de la matriz de extracción.

Geográficamente, el corpus muestra un desequilibrio estructural relevante: 142 de los 153 estudios (92,8 %) provienen de contextos fuera de América Latina, y solo 11 (7,2 %) corresponden a la región latinoamericana, distribuidos entre México (4), Argentina (2), Brasil (2),

Colombia (1), Chile (1) y Ecuador (1). La composición lingüística completa este cuadro: 149 estudios (97,4 %) están escritos íntegramente en inglés, un único estudio está íntegramente en español (Corica et al., 2025, de Argentina) y tres son bilingües o multilingües con el inglés como lengua principal. Esta subrepresentación refuerza la pertinencia regional del presente estudio: la literatura disponible para fundamentar decisiones didácticas en el contexto iberoamericano proviene, en su inmensa mayoría, de contextos culturales, curriculares e institucionales distintos al nuestro, lo que obliga a que cualquier transferencia de marcos se acompañe de un análisis crítico de su aplicabilidad contextual (Chacón-Rivadeneira et al., 2024).

Tipología de IA y herramientas reportadas

La codificación del tipo de IA muestra un claro predominio de los sistemas generativos: 69 estudios (45,1 %) abordan exclusivamente IA generativa, 66 (43,1 %) combinan dos o más tipologías, 15 (9,8 %) se centran en analítica y aprendizaje automático, 2 (1,3 %) abordan otras tipologías y un único estudio se centra exclusivamente en tutores inteligentes clásicos pre-LLM. La Tabla 2 presenta las herramientas concretas con mayor presencia en el corpus.

Tabla 2

Herramientas de IA con mayor presencia en el corpus (N = 153)

Herramienta	Corpus (n=153)	Subc. matem. (n=112)	Tipo principal
ChatGPT / GPT (3.5, 4, 4o, Trained)	57	51	Modelo de lenguaje grande (LLM)
Herramientas propias / plataformas experimentales	93	55	Variable según estudio
Gemini / Bard (Google)	1	1	LLM multimodal
GeoGebra Discovery (componente IA)	1	1	Razonamiento automatizado en geometría
Midjourney	1	1	Generación de imágenes
DALL-E (OpenAI)	1	1	Generación de imágenes
Claude (Anthropic)	1	0	LLM, interacción conversacional

Nota. La categoría «Herramientas propias» agrupa estudios que desarrollan o usan sistemas no comerciales o que no especifican una herramienta comercial concreta. Un estudio puede nombrar más de una herramienta, por lo que los conteos no son mutuamente excluyentes. Elaboración propia a partir de la matriz de extracción.

La hegemonía de ChatGPT y sus variantes (37,3 % del corpus al menos una vez; 45,5 % en el subcorpus matemática) tiene una implicación metodológica relevante: una fracción significativa de los marcos teóricos se han desarrollado con referencia implícita a las capacidades y limitaciones específicas de los modelos de lenguaje conversacionales de OpenAI. Cualquier transferencia a otras herramientas (Claude, Gemini o modelos de código abierto) requiere examen cuidadoso, particularmente en relación con los marcos del núcleo crítico-ético, donde las políticas de seguridad, los sesgos y las alucinaciones presentan perfiles diferenciados por modelo.

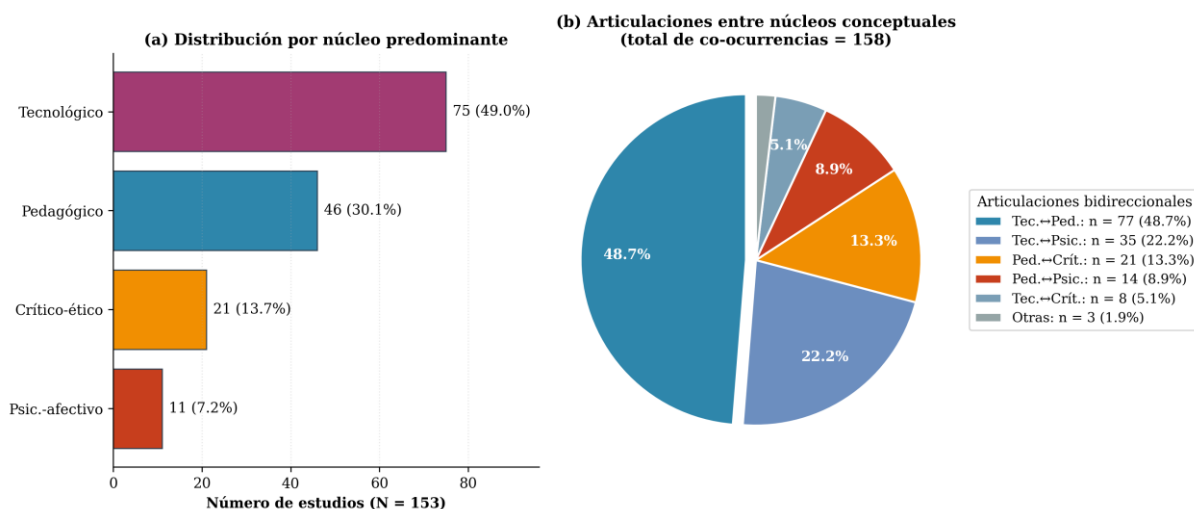
Tipología por núcleos conceptuales y articulación inter-núcleos

La clasificación de los 153 estudios según su núcleo predominante revela un perfil claramente asimétrico (Figura 3, panel a). El núcleo tecnológico agrupa el 49,0 % de los estudios ($n = 75$), seguido del pedagógico con el 30,1 % ($n = 46$), el crítico-ético con el 13,7 % ($n = 21$) y el psicológico-afectivo con el 7,2 % ($n = 11$). Esta distribución confirma cuantitativamente el diagnóstico que orientó el estudio: la producción académica del periodo 2020–2025 sigue centrada en modelos de integración tecnológica y de aceptación —particularmente TPACK, TAM y UTAUT—, mientras que los marcos que problematizan dimensiones éticas y psicológico-afectivas de la interacción con IA ocupan un lugar minoritario. La brecha entre el núcleo tecnológico y el psicológico-afectivo (41,8 puntos porcentuales) documenta un vacío que debe interpretarse como prioritario para la investigación futura.

El panel (b) de la Figura 3 sintetiza las 158 articulaciones bidireccionales entre núcleos identificadas en el corpus. La articulación entre los núcleos tecnológico y pedagógico concentra el 48,7 % del total (77 articulaciones), configurándose como el eje central del campo y reflejando la hegemonía de TPACK y sus extensiones, marcos que por diseño articulan conocimiento tecnológico, pedagógico y disciplinar (Ning et al., 2024; Priyanda et al., 2025). La articulación entre tecnológico y psicológico-afectivo alcanza el 22,2 % (35 articulaciones), sostenida principalmente por estudios que combinan TAM o UTAUT con autoeficacia o carga cognitiva. En cambio, las articulaciones que involucran al núcleo crítico-ético son claramente minoritarias: 13,3 % con el pedagógico y solo 5,1 % con el tecnológico. Esta escasez cuantifica una brecha teórica señalada cualitativamente por Gallent-Torres et al. (2023): las consideraciones éticas permanecen escindidas tanto del diseño pedagógico como del diseño tecnológico.

Figura 3

Distribución de estudios por núcleo conceptual y articulaciones entre núcleos



Nota. En el panel (a) los núcleos se asignan por el lente teórico predominante; los porcentajes suman 100 %. El panel (b) representa las 158 articulaciones bidireccionales entre pares de núcleos (combinación entre núcleo predominante y núcleo secundario de cada estudio). Elaboración propia a partir de la matriz de extracción.

Marcos teóricos más frecuentes

La codificación manual de los nombres de marcos y constructos declarados en los 153 estudios identificó más de cincuenta marcos distintos. Para manejar esta diversidad se aplicó un procedimiento de agregación por familias conceptuales; por ejemplo, la familia TPACK agrupa el TPACK original de Mishra y Koehler (2006), las extensiones hacia IA (AI-TPACK, IE-TPACK) y las variantes aplicadas a la didáctica matemática (KTMT). La Tabla 3 lista los quince marcos más frecuentes junto con su presencia específica en el subcorpus matemática.

Tabla 3

Marcos teóricos más frecuentes en el corpus (N = 153)

Marco teórico	Corpus (n)	% corpus	Subc. matem. (n / % rel.)	Núcleo
TPACK y extensiones (AI-TPACK, IE-TPACK, KTMT)	22	14,4 %	17 / 77,3 %	Tecnológico
Marcos críticos y éticos (agregado)	21	13,7 %	17 / 81,0 %	Crítico-ético
PCK — Pedagogical Content Knowledge (Shulman)	16	10,5 %	12 / 75,0 %	Pedagógico
Autoeficacia (Bandura, Usher y Pajares)	11	7,2 %	3 / 27,3 %	Psic.-afectivo
TAM — Technology Acceptance Model (Davis)	10	6,5 %	5 / 50,0 %	Tecnológico

Scaffolding / Andamiaje (Wood, Bruner y Ross)	9	5,9 %	7 / 77,8 %	Pedagógico
AI Literacy (Ng et al.; Long y Magerko)	8	5,2 %	2 / 25,0 %	Tecnológico
Enfoque instrumental (Rabardel/Trouche/Artigue)	6	3,9 %	6 / 100,0 %	Pedagógico*
Metacognición (Flavell; Schraw y Moshman)	5	3,3 %	3 / 60,0 %	Psic.-afectivo
Taxonomía de Bloom y extensiones digitales	5	3,3 %	1 / 20,0 %	Pedagógico
Pensamiento computacional (Wing; Weintrop et al.)	5	3,3 %	1 / 20,0 %	Pedagógico
ZDP / Vygotsky y perspectiva sociocultural	4	2,6 %	3 / 75,0 %	Pedagógico
PBL — Aprendizaje Basado en Problemas	4	2,6 %	3 / 75,0 %	Pedagógico
Constructivismo (piagetiano y general)	4	2,6 %	2 / 50,0 %	Pedagógico
UTAUT (Venkatesh et al.)	3	2,0 %	2 / 66,7 %	Tecnológico

Nota. La fila «Marcos críticos y éticos» agrupa los 21 estudios cuyo núcleo predominante es crítico-ético e incluye marcos nominalmente éticos (FAccT AI, integridad académica, HCAI) así como marcos analíticos de ese núcleo (SIPE-AI, SFG, Actor-Network Theory, epistemic stance). El asterisco () señala marcos con especificidad disciplinar matemática. Los porcentajes del subcorpus matemática son relativos al total de estudios que declaran ese marco. Elaboración propia a partir de la matriz de extracción.*

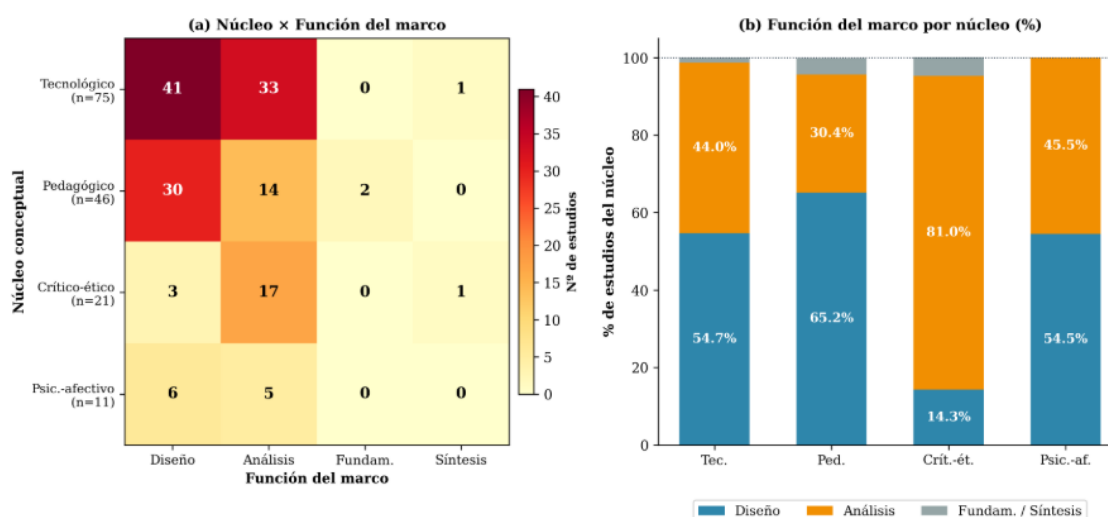
Los hallazgos merecen destacarse de este ranking. Primero, la dominancia de TPACK y sus extensiones (14,4 % del corpus) confirma que los marcos de integración tecnológico-pedagógica son la lente analítica privilegiada del campo. En el subcorpus matemática, TPACK aparece en 17 de sus 22 registros (77,3 %), lo que indica que el marco ha sido adoptado con especial intensidad por la investigación en educación matemática. Segundo, el 66,7 % del corpus (n = 102) adopta al menos un marco originado en la didáctica matemática; esta proporción se eleva al 90,2 % en el subcorpus matemática (101 de 112 estudios), lo que contradice la lectura superficial según la cual la investigación sobre IA en matemática suele apoyarse en marcos genéricos. Marcos como el enfoque instrumental (Rabardel/Trouche/Artigue), la Idoneidad Didáctica del EOS (Pochulu y Font, 2025) o PCK aplicado a matemática exhiben una presencia disciplinar fuerte y documentan una tradición didáctico-matemática consolidada que dialoga activamente con la IA.

Función del marco y nivel de generalidad

El cruce entre núcleo conceptual y función del marco ofrece uno de los hallazgos más reveladores del estudio (Figura 4). El 52,3 % de los 153 estudios utiliza el marco con función de diseño (n = 80), el 45,1 % con función de análisis (n = 69), y solo el 2,6 % lo emplea en funciones de fundamentación o síntesis. Por núcleo, las diferencias son pronunciadas: el núcleo tecnológico distribuye sus estudios entre diseño (54,7 %) y análisis (44,0 %) de manera relativamente equilibrada; el pedagógico se inclina hacia el diseño (65,2 %), coherente con la naturaleza prescriptiva de los modelos instruccionales y de la didáctica específica; el psicológico-afectivo mantiene un equilibrio similar al tecnológico (54,5 % y 45,5 % respectivamente). Pero el núcleo crítico-ético se inclina decididamente hacia el análisis (81,0 %), con solo tres estudios (14,3 %) empleándolo como marco de diseño.

Figura 4

Cruce entre núcleo conceptual y función del marco en el estudio



Nota. El panel (a) presenta las frecuencias absolutas; el panel (b) muestra los porcentajes dentro de cada núcleo. Las funciones de fundamentación y síntesis se agrupan en el panel (b) por su frecuencia marginal. Elaboración propia a partir de la matriz de extracción.

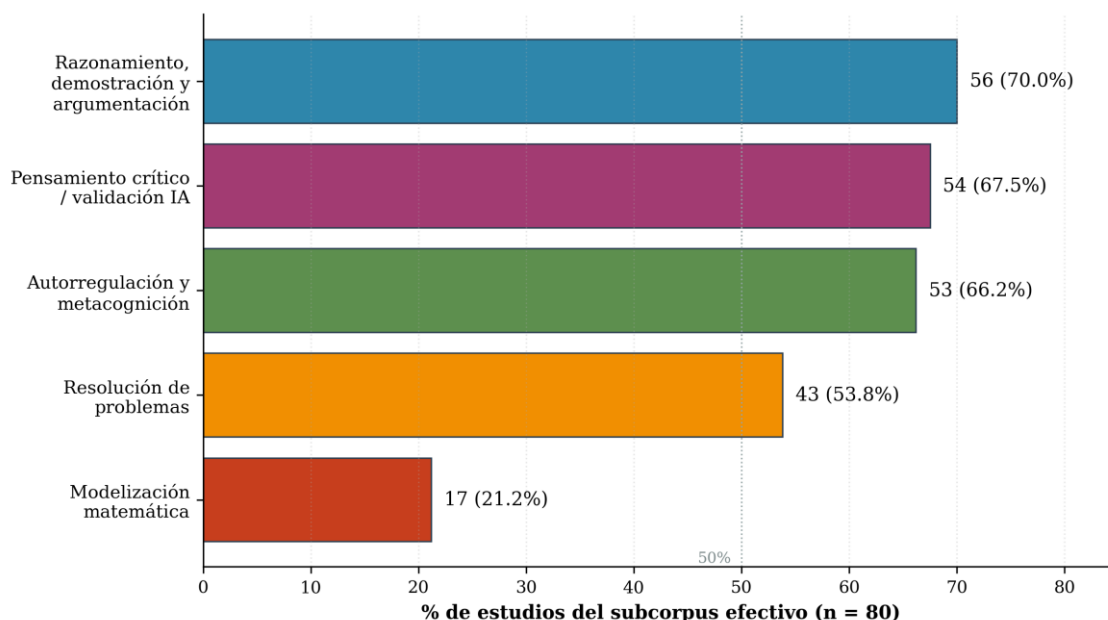
Este patrón tiene una implicación pedagógica crucial: los marcos éticos, en el corpus, rara vez se operacionalizan como criterios de diseño de la intervención, sino que tienden a emplearse como lente evaluativa sobre prácticas ya ocurridas. La brecha entre el discurso normativo ético y la prescripción didáctica concreta constituye así una línea futura relevante. La clasificación por nivel de generalidad complementa este análisis: el corpus se concentra fuertemente en el nivel meso (65,4 %; n = 100), con marcos como TPACK, TAM, Idoneidad Didáctica o los ciclos instruccionales 5E; los marcos de nivel micro representan el 28,1 % (n = 43) —donde predominan los constructos del núcleo psicológico-afectivo (63,6 % de ese núcleo es micro)— y los marcos macro solo el 6,5 % (n = 10), lo que documenta un vacío metateórico que afecta particularmente al núcleo tecnológico.

Aplicabilidad a la enseñanza de la matemática

El análisis focalizado del subcorpus efectivo ($n = 80$) permite caracterizar qué procesos matemáticos resultan efectivamente promovidos mediante la integración de IA. La Figura 5 presenta los resultados para los cinco procesos codificados en el bloque G de la matriz.

Figura 5

Procesos matemáticos promovidos por la IA en el subcorpus efectivo ($n = 80$)



Nota. Las frecuencias corresponden a la presencia (explícita o implícita) de cada proceso en el estudio. Un mismo estudio puede contribuir a varios procesos. Elaboración propia a partir del bloque G de la matriz de extracción.

Los hallazgos, ordenados de mayor a menor frecuencia, son los siguientes: el razonamiento, la demostración y la argumentación están presentes en 56 de 80 estudios (70,0 %); el pensamiento crítico orientado a validar las respuestas de la IA en 54 (67,5 %); la autorregulación y la metacognición en 53 (66,2 %); la resolución de problemas en 43 (53,8 %); y la modelización matemática solo en 17 estudios (21,2 %). Esta distribución constituye, a juicio de los autores, el hallazgo pedagógicamente más relevante de la revisión.

El dato más llamativo es la fuerte presencia del pensamiento crítico orientado a validar las respuestas de la IA (67,5 %). Este proceso, que apenas existía como categoría didáctica antes de la irrupción de la IA generativa, se ha convertido en un eje central de la actividad matemática mediada por IA. La proximidad entre este porcentaje y el del razonamiento y demostración (70,0 %) sugiere que ambos procesos están íntimamente entrelazados en el corpus: el razonamiento matemático con IA se ejerce, en buena medida, como razonamiento sobre la IA. La literatura aborda este desplazamiento bajo la figura del estudiante como «auditor» de las soluciones producidas por la IA (Pochulu y Font, 2025), del docente como «orquestador» que modula cuándo confiar en la herramienta (Acuña y Marrero, 2025), y como estrategia explícita para gestionar el riesgo de alucinaciones y demostraciones falsas plausibles (Yoon et al., 2024).

Por contraste, la modelización matemática presenta la menor cobertura (21,2 %): 62 de los 80 estudios del subcorpus efectivo no reportan este proceso. Este dato representa un vacío relevante. La IA generativa tiene el potencial teórico de apoyar procesos de modelización al conectar fenómenos reales con representaciones matemáticas complejas, pero la literatura empírica del periodo todavía no ha integrado sistemáticamente este proceso como objeto de diseño ni de análisis. Este vacío contrasta con el peso histórico que la modelización tiene en la didáctica de la matemática (PISA, ICMI, currículos nacionales) y constituye una línea futura con alto potencial didáctico.

Tabla 4

Contenidos matemáticos abordados en el subcorpus efectivo (n = 80)

Contenido matemático	n	%	Niveles predominantes
Álgebra (ecuaciones, sistemas, polinomios)	15	18,8 %	Secundaria, Superior
Geometría euclidiana y transformacional	13	16,2 %	Primaria, Secundaria
Aritmética, números y operaciones básicas	12	15,0 %	Primaria
Cálculo y precálculo	8	10,0 %	Superior (STEM)
Modelización matemática	5	6,2 %	Secundaria, Superior
Estadística y probabilidad	5	6,2 %	Secundaria, Superior
Funciones	4	5,0 %	Secundaria, Superior
Trigonometría	2	2,5 %	Secundaria, Superior
Lógica y razonamiento formal	2	2,5 %	Superior
Combinatoria	1	1,2 %	Secundaria
Fractales y geometría no euclidiana	1	1,2 %	Superior

Nota. La clasificación no es mutuamente excluyente: un estudio puede abordar más de un contenido. Los porcentajes se calculan sobre n = 80. Elaboración propia a partir de la matriz de extracción.

En términos de contenidos matemáticos trabajados (Tabla 4), el álgebra (18,8 %), la geometría (16,2 %) y la aritmética (15,0 %) concentran en conjunto el 50,0 % del subcorpus efectivo; el cálculo y el precálculo agregan otro 10,0 %, vinculado mayoritariamente a educación superior STEM. La experiencia de Acuña y Marrero (2025) en un curso semipresencial de sucesiones y series del Instituto Tecnológico de Costa Rica se inscribe en este último grupo y constituye, junto con los trabajos del TEC sobre actitudes estudiantiles (Meza-Cascante et al., 2026) e innovación didáctica con IA generativa (Ramírez-Bogantes et al., 2026), un referente local directamente relevante para la transferencia de los hallazgos del presente estudio al contexto costarricense.

Frente a la paradoja pedagógica que identifican estos hallazgos —la IA generativa resulta más asequible para tareas donde la demostración puede fallar y menos para la modelización donde podría aportar más—, los estudios que incorporan marcos específicos de la didáctica de la matemática ofrecen un camino alternativo. Particularmente relevantes son la Idoneidad Didáctica del EOS (Pochulu y Font, 2025) y el enfoque instrumental articulado con génesis instrumental, que proponen diseñar tareas donde el estudiante no use la IA para obtener la solución sino para auditar, validar o refutar la solución que la IA propone. Esta reconfiguración invierte la relación tradicional entre estudiante y herramienta: convierte las limitaciones de la IA —sus errores, sus alucinaciones, sus demostraciones plausibles pero falsas— en un recurso pedagógico central (Yoon et al., 2024).

CONCLUSIONES

La revisión exploratoria de 153 estudios publicados entre 2020 y 2025 sobre marcos teóricos de IA en educación permitió sistematizar un campo hasta ahora fragmentado y derivar conclusiones operativas para la enseñanza de la matemática. La concentración del 82,4 % del corpus en el bienio 2024–2025 confirma que se trata de una literatura emergente, reconfigurada por la irrupción masiva de los modelos de lenguaje generativos; los resultados del presente estudio deben interpretarse, en consecuencia, como una fotografía temporal de un campo en rápida transformación.

El predominio del núcleo tecnológico (49,0 %), seguido del pedagógico (30,1 %), configura el eje central del campo. La articulación entre ambos núcleos concentra el 48,7 % de las 158 co-ocurrencias identificadas y se sostiene, principalmente, en TPACK y sus extensiones hacia IA, que constituyen el marco más recurrente con el 14,4 % del corpus. Este eje, por su versatilidad y operatividad, será probablemente el andamiaje sobre el cual se construyan las propuestas de integración de IA en los próximos años. Sin embargo, el estudio documenta también que los marcos propios de la didáctica matemática —enfoque instrumental, Idoneidad Didáctica, APOS, RME— mantienen una presencia robusta en el subcorpus de matemática (90,2 % de sus estudios), lo que contradice la lectura común según la cual el campo estaría dominado por marcos genéricos.

La brecha más significativa detectada concierne al núcleo crítico-ético. Si bien alcanza el 13,7 % del corpus y muestra un crecimiento sostenido desde 2024, opera predominantemente como lente analítica (81,0 %) y rara vez como marco de diseño (14,3 %). Sus articulaciones con otros núcleos son limitadas: solo 13,3 % con el pedagógico y 5,1 % con el tecnológico. Este aislamiento relativo indica que el discurso ético sobre la IA en educación, aunque creciente, aún no se ha traducido al diseño didáctico concreto ni a las dimensiones afectivas del aprendizaje. Cerrar esta brecha constituye una tarea prioritaria para la investigación futura y para la formación docente.

En el subcorpus efectivo de enseñanza de la matemática ($n = 80$), el hallazgo central es que el razonamiento y demostración (70,0 %) y el pensamiento crítico orientado a validar la IA (67,5 %) constituyen los procesos más promovidos, con cifras casi idénticas que sugieren un entrelazamiento conceptual sustantivo: el razonamiento matemático con IA se ejerce, en buena medida, como razonamiento crítico sobre la IA. Este desplazamiento reconfigura el rol pedagógico de la herramienta: el estudiante no delega la demostración en la máquina sino que ejerce la validación lógica sobre los argumentos que la máquina produce. La modelización matemática, en contraste, aparece subrepresentada (21,2 %), lo que constituye un vacío relevante por el peso formativo que esta competencia tiene en los currículos contemporáneos.

Como recomendaciones operativas para el diseño didáctico, la mediación docente y la evaluación del aprendizaje matemático mediado por IA, el estudio sugiere cuatro orientaciones. Primera, explicitar el marco teórico en cada experiencia de integración y evitar el uso puramente nominal; los criterios de operacionalización propuestos (Creswell y Creswell, 2018; Tricco et al., 2018) ofrecen una guía adecuada. Segunda, privilegiar marcos de nivel meso como TPACK / AI-TPACK o Idoneidad Didáctica del EOS por su operatividad tanto en el diseño como en el análisis, sin renunciar a la complementariedad con constructos micro (autoeficacia, metacognición, carga cognitiva). Tercera, incorporar sistemáticamente criterios éticos al diseño didáctico, no solo a la fundamentación declarativa; la brecha documentada entre el 81,0 % de uso analítico y el 14,3 % de uso prescriptivo del núcleo crítico-ético debe cerrarse. Cuarta, reorientar las tareas con IA hacia la validación lógica y la auditoría crítica del estudiante, especialmente en dominios de alto valor formativo como la demostración y la modelización.

Entre las limitaciones del estudio conviene señalar tres. La ventana temporal 2020–2025 prioriza actualidad pero puede subrepresentar desarrollos muy recientes; la heterogeneidad terminológica del campo exigió iteraciones metodológicas; y la representación latinoamericana (7,2 % del corpus), consistente con el alcance anglófono de las bases de datos principales, obliga a cualquier transferencia de los marcos al contexto iberoamericano a acompañarse de un análisis crítico de su aplicabilidad contextual. Como líneas futuras prioritarias se identifican: la construcción de marcos híbridos que integren explícitamente la dimensión ética con la didáctica específica de la matemática; el desarrollo de instrumentos para evaluar el razonamiento estudiantil en tareas mediadas por IA generativa; y la exploración sistemática del potencial de la IA para apoyar la modelización matemática, proceso donde su contribución permanece casi inexplorada en la literatura actual.

REFERENCIAS

- Acuña, C. R. Y., y Marrero, C. M. (2025). Integración de 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>
- Arksey, H., y 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>
- 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>
- Creswell, J. W., y Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5.ª ed.). SAGE Publications.
- Gallent-Torres, C., Zapata-González, A., y Ortego-Hernando, J. L. (2023). El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa, 29(2). <https://doi.org/10.30827/relieve.v29i2.29134>
- 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.
- Mishra, P., y Koehler, M. J. (2006). Technological pedagogical content knowledge: a framework for teacher knowledge. Teachers College Record, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Ning, Y., Zhang, C., Xu, B., Zhou, Y., y Wijaya, T. T. (2024). Teachers' AI-TPACK: exploring the relationship between knowledge elements. Sustainability, 16(3), 978. <https://doi.org/10.3390/su16030978>
- 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>
- 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>
- Priyanda, R., Herman, T., Amalia, R., y Ihsan, I. R. (2025). Exploring teachers' pedagogical reasoning in mathematics education using the TPACK framework. *Frontiers in Education*, 10, 1552760. <https://doi.org/10.3389/educ.2025.1552760>
- Ramírez-Bogantes, M., Monge-Fallas, J., Borbón-Alpizar, A., Gutiérrez-Montenegro, M. V., Acuña-Chacón, R., y Meza-Cascante, L. G. (2026). Innovación didáctica con IA-G generativa en matemática universitaria: experiencias en el TEC. *Revista Tecnología en Marcha*, 39(5), 147–156.
- 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>
- Wardat, Y., Tashtoush, M. A., AlAli, R., y Jarrah, A. M. (2023). ChatGPT: a revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7), em2286. <https://doi.org/10.29333/ejmste/13272>
- 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>