JARME Journal of Authentic Research on Mathematics Education
Volume 8, No. 2, July 2026
https://doi.org/ 10.37058/jarme.v3i1.18242 e-ISSN 2655-7762

GENERATIVE AI IN MATHEMATICS EDUCATION: FROM LEARNING COMPANION TO COGNITIVE SHORTCUT

Rival Muhamad Fauzi[1]*, Angga Nurohman[2], Dicky Zaini Abdullah Al As'ari[3]
[123]Universitas Siliwangi, Jl. Siliwangi, No. 24 Kota Tasikmalaya 46115, Indonesia
E-mail: *@rivalfzie@gmail.com

ARTICLE INFO

Article history
Received: 2026-04-30
Revised: 2026-05-12
Accepted: 2026-08-22

Keywords
Generative Artificial Intelligence, Mathematics Education, Learning Companion, Cognitive Shortcut, Systematic Literature Review, Bibliometric Analysis

ABSTRACT

The emergence of generative artificial intelligence (GenAI) presents new opportunities in mathematics education through interactive support, feedback, scaffolding, and problem solving, while also creating risks when AI becomes a shortcut that replaces students' thinking processes. This study aims to map research developments on GenAI in mathematics education, identify opportunities and challenges associated with its use, and formulate directions for future research. A Systematic Literature Review (SLR) complemented by bibliometric analysis and thematic synthesis was conducted. Literature was searched through Scopus for publications from 2016 to 2025. Of 105 identified records, 45 studies met the inclusion criteria and were substantively analyzed following the PRISMA 2020 procedure. Bibliometric analysis using VOSviewer and thematic synthesis identified three major themes: AI as Learning Companion, AI as Cognitive Shortcut, and Future Research Directions. The findings reveal a shift from technology-oriented research toward pedagogical integration involving learning, teacher support, problem solving, and student–AI interaction. GenAI can strengthen learning when it supports students' thinking processes but may undermine independence when it substitutes for mathematical reasoning. These findings highlight the importance of pedagogical design, student agency, AI literacy, teacher roles, and assessment practices that preserve students' cognitive engagement.

Perkembangan generative artificial intelligence (GenAI) menghadirkan peluang baru dalam pendidikan matematika melalui dukungan interaktif, umpan balik, scaffolding, dan pemecahan masalah, tetapi juga menimbulkan risiko ketika AI digunakan sebagai jalan pintas yang menggantikan proses berpikir siswa. Penelitian ini bertujuan memetakan perkembangan penelitian GenAI dalam pendidikan matematika, mengidentifikasi peluang dan tantangan penggunaannya, serta merumuskan arah penelitian masa depan. Penelitian menggunakan Systematic Literature Review (SLR) yang dilengkapi analisis bibliometrik dan thematic synthesis. Pencarian dilakukan melalui Scopus pada publikasi 2016–2025. Dari 105 records yang teridentifikasi, 45 studi memenuhi kriteria inklusi dan dianalisis secara substantif menggunakan prosedur PRISMA 2020. Analisis bibliometrik menggunakan VOSviewer dan thematic synthesis mengidentifikasi tiga tema utama: AI as Learning Companion, AI as Cognitive Shortcut, dan Future Research Directions. Hasil menunjukkan pergeseran fokus penelitian dari orientasi *teknologi menuju integrasi pedagogis yang mencakup pembelajaran, dukungan guru, pemecahan masalah, dan interaksi siswa–AI. GenAI berpotensi memperkuat pembelajaran ketika mendukung proses berpikir, tetapi dapat mengurangi kemandirian ketika menggantikan penalaran siswa. Temuan ini menegaskan pentingnya desain pedagogis, agensi siswa, literasi AI, peran guru, dan asesmen yang menjaga keterlibatan kognitif siswa.*

How to Cite:
Fauzi, R.M., Nurohman, A., & As'ari, D.Z.A., (2026). Generative Ai In Mathematics Education: From Learning Companion To Cognitive Shortcut . ***Journal of Authentic Research on Mathematics Education, 8***(2), 53-74. https://doi.org/10.37058/jarme.v3i1.18242

1. INTRODUCTION

Generative Artificial Intelligence (GenAI) is developing rapidly and increasingly entering educational practice, including mathematics education. Its ability to generate explanations, provide examples, respond to questions interactively, and provide assistance that can be adapted to users' needs opens opportunities to support mathematics learning. In this context, GenAI has the potential to function as a *learning companion*, namely a learning partner that provides flexible support, offers scaffolding, helps students explore solution strategies, and provides feedback during the learning process. Empirical research on the use of GenAI in mathematics education indicates that large language models and ChatGPT can be utilized to support problem solving, problem-based mathematics learning, the development of learning environments, and various forms of pedagogical support for teachers and students (Liu et al., 2025; Malik et al., 2025; Li et al., 2025).

However, the ability of GenAI to generate answers quickly, completely, and convincingly also presents pedagogical concerns that cannot be overlooked. In mathematics learning, successfully obtaining a correct answer does not necessarily indicate that students understand the underlying concepts, strategies, or reasoning processes (Ganaie, 2026; Li & Bin Awang, 2026). When AI is used to support exploration, provide hints, or test strategies, its use can strengthen students' cognitive engagement (Cosentino et al., 2025; Walkington, 2025). Conversely, when students delegate most of the reasoning process to AI and merely take the final answer, the same technology may function as a *cognitive shortcut* (Cao & Abdullah, 2025; Redmond Sanogo et al., 2025). Wang et al. (2025) demonstrate a tension between GenAI as a scaffold and a crutch, particularly when AI use is directed more toward obtaining answers than developing understanding (Ganaie, 2026). Findings concerning students' interaction with generative AI teachable agents also indicate that increasing AI authority may be associated with changes in students' agency during the learning process (Xing et al., 2026). Thus, the effectiveness of GenAI in mathematics education cannot be understood solely in terms of improved performance or task-completion efficiency, but must also be examined in relation to its contribution to students' reasoning, agency, and learning independence (Cosentino et al., 2025; Ganaie, 2026; Walkington, 2025).

This issue has become increasingly important as research on AI in mathematics education has developed along increasingly diverse lines. The literature addresses not only the use of AI as a learning technology, but also support for problem solving, the development of learning environments, teachers' work, student–AI interaction, AI literacy, dependence, and issues of ethics and academic integrity (Cosentino et al., 2025; Ganaie, 2026; Shang, 2025; Walkington, 2025; Yi, 2025). Studies examining the use of GenAI for mathematical proof, for example, indicate that student–AI interaction needs to be understood not only in terms of outcomes, but also in terms of how students use and evaluate AI-generated outputs during mathematical processes (Cao & Abdullah, 2025; Creely, 2026; Ganaie, 2026; Walkington, 2025; Yoon et al., 2024). On the other hand, studies concerning dependence on AI dialogue systems indicate growing concern regarding their potential effects on students' cognitive abilities (Allison et al., 2025; Lyu et al., 2026; Walkington, 2025; Zhai et al., 2024). These developments indicate that research on GenAI in mathematics education has moved beyond the question of whether AI can be used toward more fundamental questions concerning how, for what purposes, and under what conditions AI should be used in mathematics learning (Walkington, 2025; Yi, 2025).

Nevertheless, this body of research still reveals a knowledge landscape distributed across various focuses and approaches. Some studies emphasize the potential of GenAI to improve learning support and teachers' pedagogical work, whereas others highlight dependence, the delegation of reasoning processes, changes in student agency, and the risks of unguided AI use (Allison et al., 2025; Coman et al., 2026; Cosentino et al., 2025; Ganaie, 2026; Shang, 2025). This condition indicates the need for a synthesis that does not merely inventory benefits and challenges separately, but also explains the relationships between them (Ganaie, 2026; Yi, 2025; Zaimoğlu & Dağtaş, 2025). Specifically, a more systematic understanding is still needed regarding the development of research foci on GenAI in mathematics education, the characteristics of GenAI use that support students' thinking processes, and the conditions under which AI may shift from a learning companion to a substitute for the reasoning process (Lee et al., 2026; Walkington, 2025; Yi, 2025). This need is increasingly relevant because research on GenAI is developing rapidly, while the consolidation of empirical findings and conceptual understanding of its implications for mathematics education is still ongoing (Ganaie, 2026; Lee et al., 2026; Yi, 2025).

To address this gap, this study integrates a Systematic Literature Review (SLR), bibliometric analysis, and thematic synthesis. The SLR is used to identify, select, and synthesize relevant studies; bibliometric analysis is used to map the structure and developmental trends of the research; whereas thematic synthesis is used to integrate substantive findings from the selected studies. The literature search was conducted through the Scopus database covering publications from 2016–2025. The identification process yielded 105 records, which were subsequently screened based on the inclusion and exclusion criteria, resulting in 45 studies for substantive analysis. The separation between the bibliographic corpus and the studies included in the synthesis enables this study to combine the mapping of field development with a more in-depth analysis of relevant research findings (Ganaie, 2026; Namoun et al., 2024; Zhu et al., 2026).

The novelty of this study lies in integrating the mapping of literature development with substantive synthesis to understand the position of GenAI in mathematics education. The study does not stop at mapping publication frequencies, technology use, or identifying benefits and risks separately, but connects patterns of research development with changes in the pedagogical focus emerging from the literature (Ganaie, 2026; Zhu et al., 2026). Through this approach, the study develops an understanding that the role of GenAI in mathematics education can be positioned along a continuum: on one side, as a *learning companion* that strengthens thinking processes through support, feedback, and scaffolding; on the other, as a *cognitive shortcut* when its use replaces or reduces students' engagement in the reasoning process (Lee et al., 2026; Redmond Sanogo et al., 2025; Walkington, 2025). This perspective is subsequently used to identify pedagogical implications and research directions needed to ensure that GenAI use continues to support students' agency and mathematical independence (Cosentino et al., 2025; Lee et al., 2026; Walkington, 2025).

Based on these problems and gaps, this study aims to map the development of research on GenAI in mathematics education, synthesize opportunities and challenges associated with its use in the learning process, and formulate future research directions. Specifically, this study addresses three research questions: (1) how is GenAI utilized and positioned as a *learning companion* in mathematics education; (2) what challenges emerge when the use of GenAI potentially becomes a *cognitive shortcut* and reduces students' agency and independence; and (3) what research directions are needed to optimize the pedagogical benefits of GenAI while minimizing the risks associated with its use in mathematics education (Shang, 2025; Walkington, 2025; Yi, 2025).

This study is expected to contribute to the development of AI-based mathematics education research by clarifying the conditions that enable GenAI to support thinking processes and the conditions that may cause the technology to replace students' thinking processes. Practically, the findings may serve as a basis for teachers, researchers, and learning developers in designing GenAI use that is not solely oriented toward the efficiency of obtaining answers, but also preserves students' mathematical understanding, cognitive engagement, agency, and problem-solving abilities. Thus, the use of GenAI can be directed not toward replacing students' mathematical activities, but toward strengthening a learning process in which students remain the decision-makers and primary agents of mathematical reasoning (Coman et al., 2026; Cosentino et al., 2025; Lee et al., 2026; Walkington, 2025; Yi, 2025).

2. METODE

2.1. Research Design and Literatur Search Strategy

This study used a Systematic Literature Review (SLR) complemented by bibliometric analysis and thematic synthesisto identify, map, and synthesize the development of research on Generative Artificial Intelligence (GenAI) in mathematics education. The SLR was used to identify and synthesize relevant studies based on predetermined criteria. Bibliometric analysis was used to map the structure, interconnections, and temporal development of the literature based on publication metadata, whereas thematic synthesis was used to identify and integrate substantive patterns from the findings of the included studies.

The three approaches were used as part of a single literature review design and were not intended as mixed-method research. Bibliometric analysis and thematic synthesis were analytical approaches applied to the literature corpus to provide complementary perspectives in answering the research questions. Bibliometric analysis focused on the structure and development of the literature, whereas thematic synthesis focused on the substance of the research findings.

The literature search was conducted through the Scopus database in June 2026. Scopus was selected as a bibliographic source because it provides publication metadata required for the identification and selection of studies as well as bibliometric analysis. The search was limited to publications within the 2016–2025 period to obtain a research corpus representing the development of studies on GenAI and AI in mathematics education during the specified period. This year restriction applied specifically to the literature corpus searched and analyzed, rather than to all sources used as the theoretical foundation of the study. Therefore, fundamental sources or grand theory published before 2016 could still be used when relevant to the conceptual or methodological framework of the study.

The search strategy used a combination of keywords representing relevant GenAI and AI technologies, the mathematics education context, and learning and teaching activities. The search string used was:

("generative artificial intelligence" OR "ChatGPT" OR "large language model" OR "LLM" OR "intelligent tutoring system" OR "AI-based learning") AND ("mathematics education" OR "mathematical learning" OR "math education") AND ("student" OR "learning" OR "teaching" OR "pedagogy")

The first group of terms was used to identify publications related to GenAI or relevant AI technologies; the second group was used to restrict the context to mathematics education and learning; whereas the third group was used to capture research related to students, learning, teaching, and pedagogy.

The initial search produced 105 records. All records obtained from Scopus were exported in bibliographic format for subsequent duplicate removal, screening based on eligibility criteria, data extraction, and bibliometric analysis. The exported metadata included bibliographic information such as author names, publication year, title, publication source, DOI, abstract, author keywords, indexed keywords, and document type.

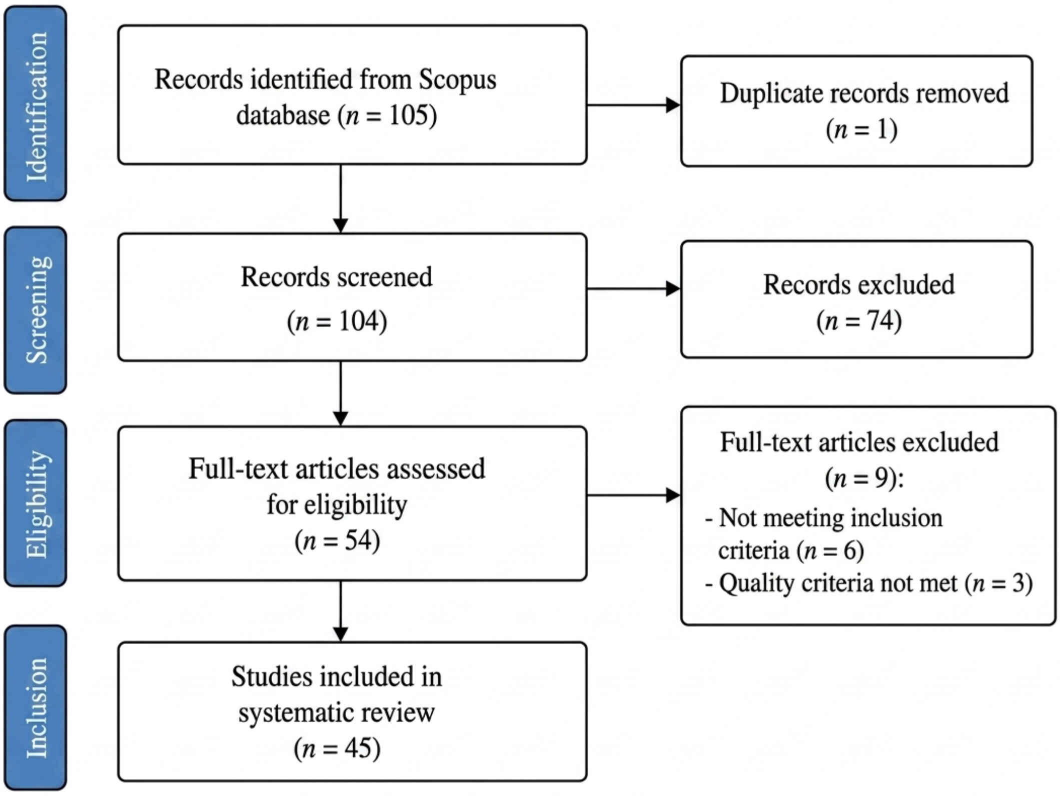

Figure 1. PRISMA 2020 flow diagram of the study selection process.

2.2. Study Selection, Data Extraction, and Analysis

2.2.1. Selection Criteria and PRISMA Procedure

Study selection was conducted systematically with reference to the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) principles to ensure that the processes of study identification, screening, eligibility assessment, and inclusion were reported transparently. PRISMA 2020 provides a reporting framework for explaining the rationale for conducting a review, the procedures employed, and the results of the selection process systematically.

The selection process consisted of several stages, namely record identification, duplicate removal, screening based on titles and abstracts, eligibility assessment based on full texts, and determination of studies meeting the inclusion criteria. Of the 105 records obtained at the identification stage, records that did not meet the research criteria were excluded during screening. Studies that passed the screening stage were subsequently assessed based on their full texts and quality criteria before being determined as studies included in the synthesis.

The inclusion and exclusion criteria were established prior to the synthesis process by considering the research focus and research questions. Studies were included if they discussed GenAI or relevant AI technologies in the context of mathematics education, learning, or teaching and made a substantive contribution to at least one research question. Studies were excluded if AI or mathematics was mentioned only peripherally, if they did not have a relevant mathematics education context, if the full text was unavailable, if they did not provide a substantive contribution to the research questions, or if they did not meet the minimum quality criteria.

Table 1. Studi Inclusion and Exclusion Criteria for Studies

Aspect	Inclusion Criteria	Exclusion Criteria
Focus	Discusses GenAI/AI in mathematics education, learning, or teaching	AI or mathematics is mentioned only peripherally
Context	Related to students, university students, teachers, preservice teachers, mathematics learning, teaching, or pedagogy	Does not have a relevant mathematics education context
Substance	Contains findings or discussion relevant to one or more RQs	Does not provide a substantive contribution to the RQs
Text	Full text is available for assessment	Full text is unavailable
Language	English	Languages other than English
Publication Type	Scientific publications that make a substantive contribution to the research focus	Editorials, non-scientific documents, or publications without a substantive contribution
Quality	Meets the minimum quality criteria	Does not meet the minimum quality criteria

2.2.2. Data Extraction

Data from the 45 included studies were systematically extracted using a structured data extraction format. The information collected included publication year, publication type, educational level, research design, AI technology or platform used, research focus, main findings, and research implications relevant to the research questions.

Data extraction was conducted in two forms. First, bibliographic data and publication characteristics were used to describe the profile of the research corpus, including the distribution of publications by year, publication type, educational level, research design, AI technology, and research focus. Second, substantive data consisting of research focus, main findings, and study implications were used as the basis for conducting the thematic synthesis.

The distinction between the 105 records and 45 studies reflects two stages in the research process. A total of 105 records represented the initial identification results from the Scopus search and served as the basis for the selection process and bibliographic mapping. Following screening, full-text assessment, and application of the eligibility criteria, 45 studies met the requirements for inclusion in the substantive synthesis. Thus, the number 105 represents the records at the identification stage, whereas the number 45 represents the studies included in the final SLR corpus.

2.2.3. Thematics Synthesis

The substantive findings from the 45 included studies were analyzed using a thematic synthesis approach. This approach was used to identify patterns of findings across studies and develop a more integrated interpretation of findings from different studies. The synthesis process followed the principle of developing themes progressively through coding of findings, grouping codes with conceptual similarities, development of descriptive themes, and formation of analytical themes. The thematic synthesis approach developed by Thomas and Harden emphasizes the systematic coding of findings to generate themes capable of explaining patterns across studies.

The first stage involved reading and examining the main findings of each study relevant to the research questions. Substantive information was then coded based on the meaning and focus of the findings. The second stage involved comparing codes with conceptual similarities and grouping them into broader categories. The third stage generated descriptive themes representing patterns of findings in the studies analyzed. Subsequently, relationships among themes were examined to generate analytical themes that explained the broader meaning of those research patterns.

The synthesis process was guided by three research questions focusing on opportunities for GenAI use, challenges associated with its use, and future research directions. However, the themes were not mechanically determined based on the structure of the research questions. Themes were developed from patterns of findings emerging from the studies and subsequently interpreted based on their relationships with the research questions.

The synthesis process resulted in three main themes:

1. AI as Learning Companion, representing the use of GenAI as a companion that provides support for the learning process, problem solving, feedback, explanations, and more adaptive learning;
2. AI as Cognitive Shortcut, representing the use of GenAI that potentially replaces part of the reasoning process, increases dependence, or reduces students' agency and independence; and
3. Future Research Directions, representing the need for further research concerning longitudinal evidence, teacher readiness, AI literacy, ethics, assessment, and the development of pedagogical frameworks.

These three themes were subsequently used as the basis for the synthesis in the Results and Discussion section, while the details of study characteristics and bibliometric patterns were presented separately to maintain the distinction between literature mapping results and substantive synthesis results.

2.2.4. Bibliometric Analysis

Bibliometric analysis was conducted using VOSviewer version 1.6.20 on publication metadata obtained from Scopus. The analysis was used to map the conceptual structure, interconnections among keywords, and temporal development of research on GenAI in mathematics education. VOSviewer was used to construct and visualize bibliometric maps based on relationships among bibliographic elements.

The primary analysis used keyword co-occurrence analysis to identify keywords that appeared simultaneously in publications. Keywords with very low occurrence frequencies were excluded from the mapping to reduce network fragmentation and retain keywords with adequate representation in the corpus. The minimum occurrence was set at 5 occurrences.

After the threshold was applied, the keywords meeting the criteria were visualized in the form of a network visualization. In this visualization, the size of the node represents the frequency of keyword occurrence, whereas the lines (links) indicate co-occurrence relationships among keywords. The strength of the relationship is represented by the thickness of the links, whereas the proximity of keyword positions indicates their relative degree of relatedness within the network. Keywords with stronger patterns of interconnection were grouped by VOSviewer into clusters.

In addition to network visualization, this study used overlay visualization to identify the temporal development of keywords. The colors in the visualization represent the average publication year associated with the occurrence of keywords. Thus, the visualization was used to identify trends in changes in research focus from topics that emerged earlier toward more recent topics.

Bibliometric analysis generated information used to address the aspect of mapping the development of the field, whereas substantive interpretation of the pedagogical meaning of these patterns was conducted through thematic synthesis. This separation of analytical functions was implemented to prevent bibliometric interpretations from being treated as direct evidence of the pedagogical effectiveness of GenAI.

2.2.5. Integration of Analytical Result

Bibliometric analysis and thematic synthesis were integrated at the interpretation stage while maintaining their respective analytical functions. Bibliometric analysis was used to demonstrate the structure, interconnections, and temporal development of the research, whereas thematic synthesis was used to explain substantive patterns identified in the included studies.

The results of keyword co-occurrence and overlay visualization were compared with the themes obtained through thematic synthesis to identify relationships between the development of research focuses and the substance of research findings. The integration was conducted at the level of interpretation, rather than by combining the two types of data into a single statistical analysis. This approach enabled the study to explain how increasing research attention to AI technology, learning systems, teacher support, and personalized learning was related to the emergence of attention to the function of GenAI as a learning companion, the risks of its use as a cognitive shortcut, and the need for future research.

Thus, the integration of the two approaches produced three layers of information: (1) the structure and development of the field based on publication metadata, (2) the characteristics and patterns of findings from the included studies, and (3) the conceptual synthesis concerning changes in the function of GenAI in mathematics education. These three layers provided the basis for interpreting the research findings in the subsequent section.

3. RESULT AND DISCUSSION

3.1. Results

This section presents the study selection results, characteristics of the corpus, publication trends, bibliometric analysis, evolution of research themes, and thematic synthesis. The presentation focuses on the data and patterns obtained from the research corpus, while interpretations regarding the meaning of the findings, their relationship to theory and previous research, and their implications are presented in the discussion section.

3.1.1. Characteristics of Included Studies

A total of 45 studies that met the inclusion criteria were further analyzed based on publication characteristics, education level, research design, AI technology used, and research focus. The complete characteristics of the included studies are presented in Table 2.

Table 2. Characteristics of Included Studies

Category	Sub-Category	Frequency n	Percentage %
Publication Years	2023	13	28.9
	2024	20	44.4
	2025(Jan-Apr)	12	26.7
Document Type	Journal Article	33	73.3
	Conference Paper	9	20.0
	Review Article	3	6.7
Education Level	Primary School	8	17.8
	Secondary School	16	35.6
	Higher Education	18	40.0
	Mixed/ Unspecified	3	6.7
Research Design	Quantitative	18	40.0
	Qualitative	9	20.0
	Mixed Methods	12	26.7
	Design Based/ Development	6	13.3
AI Technology Used	Chat GPT	28	62.2
	Other LLMS (Claude, Gemini, Copilot, Llama, etc.)	7	15.6
	AI Tools/ Platforms (e.g., Khanmigo, MathGPT)	6	13.3
	Others (RL, ML models, etc.)	4	8.9
Main Focus	Learning & Achievement	18	40.0
	Teaching & Teacher Support	10	22.2
	Problem Solving & Reasoning	9	20.0
	Attitude & Perceptions	5	11.1
	Ethics & Policy	3	6.7

3.1.2. Publication Trend

The distribution of publications by year was analyzed to describe the temporal development of the studies included in the final corpus. The publication distribution is presented in Figure 2.

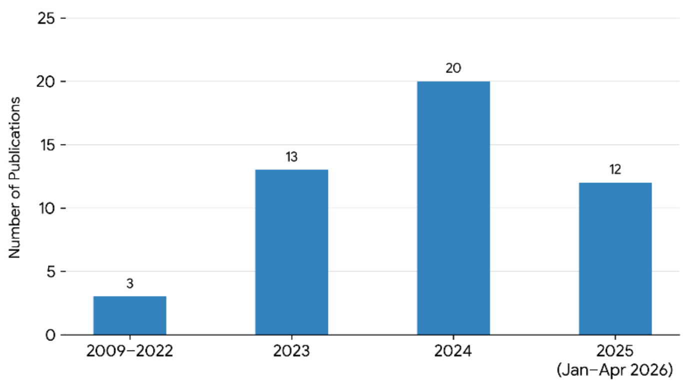

Figure 2. Annual Publication Trend of Studies on Generative AI in Mathematics Education.

3.1.3. Bibliometric Analysis

Bibliometric analysis was conducted on the publication metadata to obtain the structure of keyword relationships and the temporal development of research themes. The analysis included keyword co-occurrence networks, cluster grouping, and theme evolution.

3.1.3.1. Keyword Co-occurence Network

The results of the keyword co-occurrence analysis were visualized using overlay visualization to show the relationships among keywords within the research corpus.

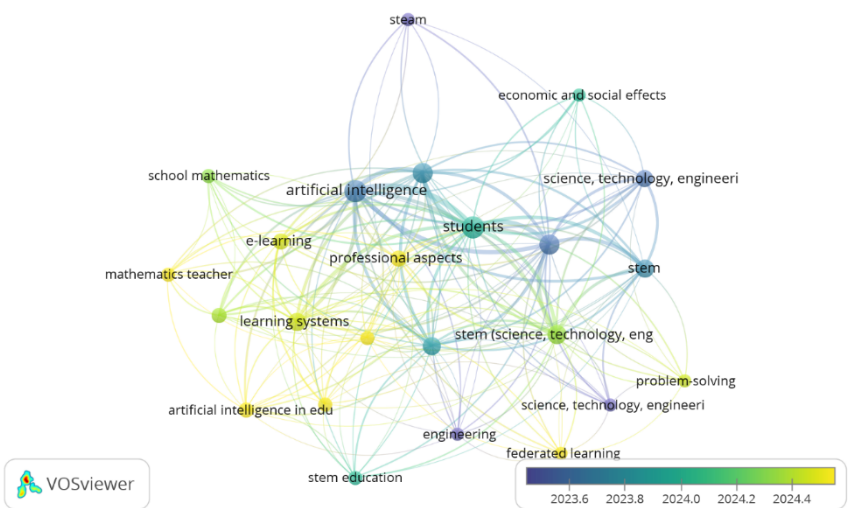

Figure 3. Overlay Visualization of Keyword Co-occurrence in Generative AI Research on Mathematics Education

The keyword co-occurrence visualization in Figure 4 shows the relational structure among the main terms in the corpus. The position of the nodes indicates the relational proximity among keywords, while the colors in the overlay visualization indicate the temporal tendency of the occurrence of terms in publications. Thus, the visualization presents not only the relationships among topics but also changes in research focus over time.

The network structure shows that research on GenAI in mathematics education is formed through relationships among technological dimensions, learning, students, teachers, and STEM contexts. This pattern provides the basis for the keyword cluster analysis in the following section.

3.1.3.1. Keyword Cluster Anaylisis

The keyword co-occurrence analysis produced five clusters based on the relationships among keywords. Details of the clusters, main keywords, descriptive interpretations based on keyword composition, and the number of items are presented in Table 3.

Table 3. Keyword Co-occurrence Clusters in Generative AI Research on Mathematics Education

Cluster	Main Keywords	Interpretation	Number of Item
1	Students; learning; motivation; attitude; achievement; problem-solving	Learning outcomes, affective factors, and problem-solving in mathematics education with GenAI	10
2	Artificial Intelligence;; chatgpt; llm; generative ai; ai tools; adaptive learning	AI technologies and their applications in learning and teaching	9
3	Stem; steam; engineering; computer science; algorithm	Technology-oriented, engineering, and STEM integration	7
4	Teaxher; professional; development; teaching support; pedagogical; curriculum	Teacher support, pedagogical implementation and professional aspects	7
5	Ethics; privacy; policy; academic integrity; risks	Ethical, policy, and socio-technical considerations	5

3.1.3.2. Evolution of Research Themes

Table 4. Keyword Co-occurrence Clusters in Generative AI Research on Mathematics Education

Cluster	Main Keywords	Interpretation	Number of Items
1	Students; learning; motivation; attitude; achievement; problem-solving	Learning outcomes, affective factors, and problem-solving in mathematics education with GenAI	10
2	Artificial Intelligence; ChatGPT; LLM; generative AI; AI tools; adaptive learning	AI technologies and their applications in learning and teaching	9

Cluster	Main Keywords	Interpretation	Number of Items
3	STEM; STEAM; engineering; computer science; algorithm	Technology-oriented, engineering, and STEM integration	7
4	Teacher; professional development; teaching support; pedagogical; curriculum	Teacher support, pedagogical implementation, and professional aspects	7
5	Ethics; privacy; policy; academic integrity; risks	Ethical, policy, and socio-technical considerations	5

3.1.3.3. Evolution of Research Themes

The temporal development of keywords was further analyzed to identify changes in research focus within the corpus. The results of the thematic evolution analysis are presented in Figure 4.

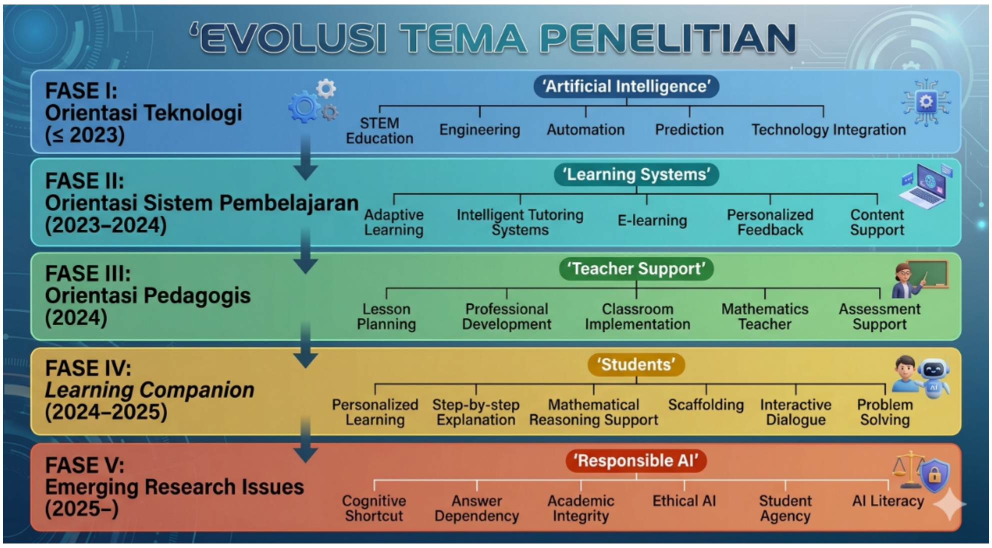

Figure 4. Thematic Evolution of Generative AI Research in Mathematics Education

3.1.5 Thematic Synthesis

Thematic synthesis was conducted on the substantive findings from the 45 included studies. The synthesis process generated themes based on the grouping of patterns that emerged across studies. Details of the themes, focus of findings, and substance of the studies supporting each theme are presented in Table 5.

Table 5. Thematic Synthesis of Generative AI Research in Mathematics Education

Theme	Focus of Findings	Main Substance
GenAI as Learning Support	Support for the mathematics learning process	GenAI is used to provide explanations, examples, feedback, and learning assistance
GenAI for Problem Solving and Reasoning	Mathematical problem solving and reasoning	GenAI is used in the problem-solving process, strategy exploration, and support for reasoning
GenAI for Teacher Support	Support for teachers' work and pedagogical practices	GenAI is utilized for lesson planning, material development, and support for pedagogical tasks

Theme	Focus of Findings	Main Substance
User Perceptions and Experiences	User perceptions, attitudes, experiences, and acceptance	Studies examine user experiences, perceptions of benefits and limitations, and acceptance of GenAI use
Ethics, Academic Integrity, and AI Literacy	Risks, ethics, academic integrity, and AI literacy	Studies highlight issues of responsible AI use, academic integrity, risks of dependence, and the need for AI literacy

3.2. DISCUSSION

3.2.1. Changes in Research Focus: From Technology Orientation toward Pedagogical Integration

The development of GenAI research in mathematics education shows a shift in attention from technological capabilities toward the pedagogical conditions that enable such technology to be used meaningfully. This shift indicates that research questions no longer stop at the capabilities of AI to generate responses or complete tasks, but are increasingly directed toward the relationships between technology, learning activities, teacher roles, and students' mathematical activities. This tendency is consistent with studies that position GenAI in mathematical problem solving, tutoring, feedback provision, personalized learning, and support for teachers' work (Walkington, 2025; Pepin et al., 2025; Yoon & Kwon, 2024; Li et al., 2025). Thus, the development of this field demonstrates a conceptual shift from technology-centered research toward pedagogically situated research (Walkington, 2025; Zaimoğlu & Dağtaş, 2025; "日本感情心理学会第26回大会プログラム," 2018).

This shift is related to the characteristics of GenAI that enable interactions to take place in a dialogic and iterative manner. Unlike the use of technologies that primarily generate outputs based on predetermined procedures, generative models allow users to request alternative explanations, ask follow-up questions, modify responses, and explore solution strategies. In mathematics education, these characteristics open possibilities for AI to be positioned as part of the learning process rather than merely as a computational tool. Liu et al. (2025) showed that a ChatGPT-based learning environment can be used to support mathematical problem solving, while Malik et al. (2025) demonstrated the potential of LLMs to provide curriculum scaffolds for teachers. These findings indicate that the pedagogical value of GenAI does not automatically reside in the technology itself, but emerges through the relationship between technological capabilities and the design of learning interactions ("<title/>," 1995; Walkington, 2025; Yi, 2025a).

The implication is that the successful integration of GenAI in mathematics education is not sufficiently determined by how accurate or complete the responses generated by the system are. More importantly, it is whether the use of such technology maintains the mathematical activities that constitute the learning objectives. In this context, the ability of GenAI to provide increasingly complex responses actually increases the need for pedagogical designs that explicitly determine the function of AI in the learning process. Thus, the development of research identified in this review can be understood as a shift from the question "what can GenAI do?" toward "how can GenAI's capabilities be positioned within the mathematics learning process without replacing students' cognitive activities?" (Walkington, 2025; Yi, 2025a)

3.2.2. GenAI as Larning Companion: When AI Supports Thinking Processes

One of the main substantive findings of this study is the emergence of GenAI as a learning companion. However, this companion function needs to be distinguished from the function of AI as a provider of answers. The value of a learning companion lies in its ability to provide support that enables students to remain engaged in the process of constructing understanding. Support in the form of step-by-step explanations, guiding questions, alternative examples, feedback, and opportunities for dialogue can help students gradually develop solution strategies. The findings of Liu et al. (2025) regarding the ChatGPT-MPS environment and Malik et al. (2025) regarding curriculum scaffolds support the understanding that GenAI can provide pedagogical support when system responses are directed toward the needs of the learning process (" and with empirical demonstrations that curriculum scaffolding and problem-solving guidance can be enhanced by GenAI-enabled interactions ("<title/>," 1995; Yi, 2025a).

From the perspective of mathematics learning, this function is important because mathematical understanding is not only related to the ability to obtain a correct result. Students need to understand relationships among concepts, the reasons underlying a procedure, and the conditions under which a strategy can or cannot be applied. Therefore, interaction with GenAI becomes pedagogical when system responses encourage students to interpret, evaluate, and use the information provided. Dilling and Herrmann (2024) showed that students' interaction with ChatGPT in mathematical reasoning and proof is influenced by their understanding of how AI works, while Yoon et al. (2024) demonstrated that the use of GenAI in proof is also related to students' conceptions of proof and ethical considerations. These findings indicate that the benefits of AI as a companion cannot be separated from users' capacity to question and evaluate the responses generated. These claims are reinforced by contemporary work on AI literacy and ethical considerations in GenAI-enabled learning environments (Demir, 2026; Dilling & Herrmann, 2024; Pepin et al., 2025) , which emphasize critical engagement with AI outputs as essential to meaningful mathematical reasoning.

Thus, a learning companion in the context of this study is more appropriately understood as a pedagogical function rather than an inherent characteristic of an AI platform. The same system can function as a companion when it provides support that maintains students' cognitive engagement, but it can become a substitute when it provides complete solutions that eliminate the need for students to think. Deng et al. (2025) further emphasize this caution by showing that the benefits of ChatGPT for academic performance and affective-motivational aspects do not by themselves demonstrate the occurrence of deeper learning. Therefore, the success of GenAI as a learning companion needs to be assessed not only based on performance when AI is available, but also based on the quality of understanding, reasoning, and students' abilities after such support is reduced or removed (Ahadzadeh et al., 2024; Pörn et al., 2024; Su, Hsu, & González Campos, 2026).

3.2.3. From Learning Companion to Cognitive Shortcut: Why Does Dependency Risk Emerge?

The findings of this study also reveal the opposite side of the learning companion function, namely GenAI as a cognitive shortcut. These two functions do not represent two different types of technology, but rather two pedagogical consequences that may emerge from the use of the same technology. The difference primarily lies in how the cognitive load is distributed between the student and the system. When AI is used to provide hints, pose questions, or help students check their strategies, most of the cognitive activity remains with the student. Conversely, when AI is used to generate complete solutions that are directly used as answers, some of the activity is transferred to the system. The tension between scaffold and crutch reported by (Wang et al., 2025) illustrates this mechanism, while Zhai et al. (2024) demonstrate the risks associated with excessive reliance on AI dialogue systems (Akinwumi et al., 2026; Su, Hsu, & Campos, 2026; Wermann et al., 2026; Zaimoğlu & Dağtaş, 2025).

This issue is particularly relevant in mathematics because problem solving is an activity that requires a series of cognitive decisions. Students not only need to obtain a result, but also identify relevant information, select representations, determine strategies, carry out procedures, check the reasonableness of the result, and provide justification. If the entire process can be transferred to GenAI, successful task completion may no longer fully represent students' mathematical abilities. Yoon et al. (2024) showed that students may treat GenAI responses as authoritative sources in proof activities, while Dilling and Herrmann (2024) demonstrated that limited understanding of LLMs can influence how students interact with the technology. Thus, the risk of a cognitive shortcut does not simply arise from the existence of AI, but from the use of AI that reduces students' opportunities to engage in mathematical activities that should constitute the learning objectives (Dilling & Herrmann, 2024; Pepin et al., 2025; Walkington, 2025).

Experimental evidence from Bastani et al. (2025) provides further explanation of this issue. The use of GPT-4 can improve student performance while AI is available, but unguarded use may be associated with lower performance when access to AI is discontinued. Conversely, the use of AI that provides hints rather than solutions can reduce such consequences. These findings clarify the distinction between AI-assisted learning and AI-substituted performance. In the context of this study, this distinction is important because it demonstrates that improved performance during interaction with GenAI cannot immediately be treated as evidence of improved independent mathematical ability (Ahadzadeh et al., 2024; Liu et al., 2025).

3.2.4. Student Agency as the Distinguishing Factor between Assistance and Replacement

The issue of cognitive shortcuts subsequently leads to the issue of student agency. When GenAI is used in learning, students still need to maintain control over key decisions in the learning process: when assistance is needed, what form of assistance is appropriate, whether an AI response is acceptable, and how the response should be verified. Thus, agency does not mean rejecting the use of AI, but ensuring that AI remains a source of support controlled by students. The findings of Xing et al. (2025), Zhang et al. (2025), and

Yoon et al. (2024) support the importance of maintaining human control in interactions with GenAI (Ganaie, 2026; Li & Bin Awang, 2026; Walkington, 2025).

This perspective explains why learning companion and cognitive shortcut cannot be distinguished solely based on the type of technology. Their pedagogical functions are shaped by the interaction among task design, patterns of use, AI literacy, and the demands of mathematical activities. AI can expand students' capacity to explore a concept when students continue to evaluate and reconstruct the responses they receive. Conversely, AI can reduce agency when students accept system outputs as final answers without verification or justification. Therefore, the findings of this study indicate that student agency is an important mechanism distinguishing AI assistance from the replacement of students' cognitive activities (Cao & Abdullah, 2025; Cosentino et al., 2025; Redmond-Sanogo et al., 2025).

These findings also help explain why research on GenAI can produce seemingly different conclusions when only performance measures are used. Liu et al. (2025) and Malik et al. (2025) demonstrate the benefits of GenAI when the technology is positioned within a pedagogical design that provides process-oriented support, whereas Wang et al. (2025) and Bastani et al. (2025) demonstrate the risks when AI enables students to replace cognitive effort with system outputs. Thus, research questions concerning the effectiveness of GenAI need to consider the conditions of use and the distribution of cognitive activity, rather than merely comparing learning outcomes between groups that use and do not use AI (Allison et al., 2025; Chegini, 2026; Haidar & Tassis, 2025; Yi, 2025).

3.2.5. Implication for Teacher, Instructional Design, and Mathematics Assessment

The findings of this study indicate that GenAI integration requires pedagogical designs that explicitly establish the function of technology within mathematical activities. Teachers cannot simply provide access to AI, but need to determine forms of interaction that support learning objectives. For example, AI can be directed to provide hints, ask students to explain the reasons behind a particular step, compare strategies, or identify errors before providing a complete solution. Such an approach positions AI as a source of scaffolding while maintaining students' cognitive responsibility. Malik et al. (2025) emphasize the importance of curricular context and expert knowledge in the use of LLMs, while Thuy et al. (2025) demonstrate the relevance of AI use to self-directed learning, autonomy, and metacognitive engagement among pre-service mathematics teachers (Allison et al., 2025; Walkington, 2025; Zhang et al., 2025). These perspectives align with recent syntheses that underscore the need to frame GenAI within curriculum design and teacher expertise to preserve mathematical agency (Chen & Alibakhshi, 2026; Lee et al., 2026)

These implications also change how the role of teachers is understood. The presence of GenAI does not automatically reduce the need for teachers because the quality of learning remains dependent on pedagogical decisions concerning objectives, tasks, forms of assistance, and evaluation of AI responses. In GenAI-based learning environments, teachers instead play an important role in designing boundaries for AI use and ensuring that interactions with technology remain aligned with mathematical objectives. Thus, the role of teachers increasingly shifts from merely providing information toward designing learning environments, guiding human–AI interactions, and evaluating the quality of students' mathematical processes (Coman et al., 2026; Desvaux et al., 2026; Webb & Galamba, 2026).

The same consequence applies to assessment. When students can generate answers with the assistance of GenAI, assessments that evaluate only final products increasingly face difficulty distinguishing students' mathematical abilities from their ability to use AI. Therefore, assessment needs to place greater attention on the process through which answers are produced. Assessment forms may include explanations of strategies, justifications, evaluation of AI responses, reflection on errors, comparison of several approaches, and measurement of problem-solving abilities after AI assistance has been removed. The findings of Bastani et al. (2025) regarding differences in performance when AI is available and unavailable, as well as the findings of Deng et al. (2025) regarding the limitations of using outputs as indicators of learning, reinforce this need. Thus, assessment in GenAI environments needs to increasingly emphasize reasoning, verification, transfer, and independent problem solving (Lyu et al., 2026; Moghe et al., 2026; Tang, 2024).

3.2.6. Future Research Directions: From AI Effectiveness toward Conditions for Effective Use

The findings of this study indicate that the GenAI research agenda in mathematics education needs to move beyond questions concerning whether AI improves learning outcomes. More substantive questions concern under what conditions, through what mechanisms, and with what interaction designs GenAI produces learning that can persist when AI assistance is no longer available. This need arises because the effectiveness of GenAI is highly likely to be influenced by task characteristics, forms of assistance, users' literacy levels, instructional design, and how learning outcomes are measured. Deng et al. (2025) emphasize the importance of research on more sustainable impacts, while Bastani et al. (2025) demonstrate that improved performance during AI use does not necessarily indicate improved independent ability.

Therefore, future research needs to place greater attention on designs that enable the distinction between performance with AI assistance and mathematical ability without AI assistance. Longitudinal designs, comparisons of AI-on/AI-off conditions, measurement of transfer, and evaluation of retention and independence can provide stronger evidence regarding whether GenAI benefits actually result in learning. Such approaches can also help explain when scaffolding becomes dependency and what factors maintain students' cognitive engagement.

In addition to research design, AI literacy needs to be positioned as part of the research agenda in mathematics education. Dilling and Herrmann (2024) demonstrate limitations in students' understanding of how LLMs work, while Zhang et al. (2025) link AI literacy, AI trust, and AI dependency with the competencies of pre-service mathematics teachers. These findings indicate that the ability to use AI productively is not only related to the ability to operate the technology, but also to the ability to understand its limitations, evaluate its outputs, and determine when AI use is appropriate for learning objectives.

Therefore, future research needs to connect AI literacy, mathematical literacy, pedagogical design, and student agency in a more integrated manner.

3.2.7. Conceptual Synthesis: Maintaining GenAI as a Companion, Not a Replacement

When all research findings are considered in an integrated manner, GenAI in mathematics education should not be positioned in binary terms as a technology that is "good" or "bad" for learning. Its pedagogical value depends on how the technology is positioned within the relationship among students, teachers, mathematical tasks, and the assessment process. When GenAI is used to provide scaffolding, open alternative strategies, provide feedback, and encourage reflection, the technology can expand students' opportunities to engage in mathematical activities. Conversely, when AI provides solutions that replace the processes of identification, reasoning, verification, and justification, the technology has the potential to become a cognitive shortcut.

Thus, the findings of this study lead to one main principle: the integration of GenAI in mathematics education should be measured based on what students continue to do after AI provides assistance, rather than only based on what AI can produce for students. This principle unifies the bibliometric findings and thematic synthesis in this study. The shift in research focus toward pedagogical integration demonstrates increasing attention to the context of AI use, while the thematic synthesis shows that the benefits and risks of GenAI center on the distribution of cognitive activity between humans and systems. Within this framework, learning companion and cognitive shortcut represent two possible pedagogical functions determined by the design of use, rather than two inherent properties of the technology.

This perspective provides implications for the direction of GenAI-based mathematics education development. The purpose of technology integration is not merely to increase the efficiency of task completion, but to maintain the fundamental functions of mathematics learning: understanding concepts, reasoning, problem solving, verification, providing justification, and developing independent thinking. Thus, GenAI is most productive when it expands students' capacity to perform these activities, rather than when it takes over these activities.

CONCLUSION

This study shows that research on Generative Artificial Intelligence (GenAI) in mathematics education has developed from a focus on technological capabilities toward issues concerning pedagogical functions, thinking processes, and student agency. The integration of systematic literature review, bibliometric analysis, and thematic synthesis shows that GenAI has the potential to function as a learning companion through scaffolding, feedback, explanations, strategy exploration, and problem-solving support. However, the direct use of GenAI to generate solutions may turn it into a cognitive shortcut that reduces students' engagement in reasoning, verification, justification, and independent construction of understanding.

The effectiveness of GenAI therefore depends on how the technology is positioned within instructional design and how cognitive activities are distributed between students and AI. GenAI can extend thinking processes when it functions as a support, but it may replace them when students delegate mathematical decision-making to AI. Therefore, instructional and assessment designs need to preserve students' reasoning processes through the use of scaffolding, guiding questions, feedback, and strategy exploration, accompanied by the strengthening of AI literacy, mathematical literacy, metacognition, and learning independence.

Future research needs to examine not only whether GenAI is effective, but under what conditions and through what pedagogical mechanisms GenAI produces learning that persists without AI assistance. Greater attention should be given to transfer, retention, independent problem solving, student agency, and differences between performance with AI assistance and abilities demonstrated independently through longitudinal designs, comparisons between AI-assisted and AI-free conditions, and examinations of various forms of scaffolding. Overall, GenAI should be positioned as a companion to mathematical thinking processes, rather than a replacement for students' thinking activities.

REFERENCESS

Adams, S. S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., Hall, J., Samsonovich, A. V., Scheutz, M., Schlesinger, M., Shapiro, S. C., & Sowa, J. F. (2012). Mapping the landscape of human-level artificial general intelligence. *AI Magazine, 33*(1), 25–41. https://doi.org/10.1609/aimag.v33i1.2322

Al Sharab, J. F., Matuk, A. S., Ali, M., & Chen, X. (2024). WIP: Exploring mathematical proficiency in AI-driven engineering environments: A survey-based study of current workforce needs. *Proceedings - Frontiers in Education Conference, FIE*. https://doi.org/10.1109/FIE61694.2024.10893062

Almoubayyed, H., Bastoni, R., Berman, S. R., Galasso, S., Jensen, M., Lester, L., Murphy, A., Swartz, M., Weldon, K., Fancsali, S. E., Gropen, J., & Ritter, S. (2023). Rewriting math word problems to improve learning outcomes for emerging readers: A randomized field trial in Carnegie Learning's MATHia. *Communications in Computer and Information Science, 1831 CCIS*, 200–205. https://doi.org/10.1007/978-3-031-36336-8_30

Anton, J., Cosentino, G., Sharma, K., Gelsomini, M., Mok, M., Giannakos, M., & Abrahamson, D. (2025). The human condition: Modal and interactive advantages of teacher over AI feedback on children's mathematical performance. *Proceedings - 24th Annual ACM Interaction Design and Children Conference, IDC 2025*, 183–203. https://doi.org/10.1145/3713043.3728863

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

del Olmo-Muñoz, J., González-Calero, J. A., Diago, P. D., Arnau, D., & Arevalillo-Herráez, M. (2023). Intelligent tutoring systems for word problem solving in COVID-19 days:

Could they have been (part of) the solution? *ZDM–Mathematics Education, 55*(1), 35–48. https://doi.org/10.1007/s11858-022-01396-w

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Dilling, F., & Herrmann, M. (2024). Using large language models to support pre-service teachers mathematical reasoning—an exploratory study on ChatGPT as an instrument for creating mathematical proofs in geometry. *Frontiers in Artificial Intelligence, 7*, 1460337. https://doi.org/10.3389/frai.2024.1460337

Drushlyak, M., Lukashova, T., Sabadosh, Y., Melnikov, I., & Semenikhina, O. (2024). Using ChatGPT for the development of critical thinking in youth: Example of inequality proof. *2024 47th ICT and Electronics Convention (MIPRO) Proceedings*, 334–339. https://doi.org/10.1109/MIPRO60963.2024.10569759

Li, D., Osman, S., Alhassora, N. S. A., Kumar, J. A., & Husain, S. K. S. (2025). Empowering K-12 mathematics teachers with artificial intelligence: A systematic review of insights, applications, and challenges. *IEEE Access, 13*, 117198–117209. https://doi.org/10.1109/ACCESS.2025.3585784

Liu, J., Sun, D., Sun, J., Wang, J., & Yu, P. L. H. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction. *Computers and Education: Artificial Intelligence, 9*, 100438. https://doi.org/10.1016/j.caeai.2025.100438

Malik, R., Abdi, D., Wang, R., & Demszky, D. (2025). Scaffolding middle school mathematics curricula with large language models. *British Journal of Educational Technology, 56*(3), 999–1027. https://doi.org/10.1111/bjet.13571

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., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education, 11*, 9–41. https://doi.org/10.1007/s40751-025-00172-1

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*, 45. https://doi.org/10.1186/1471-2288-8-45

Thuy, P. T. D., Van, N. T., Trang, N. H., Thuong, N. M., Hien, V. T. T., & Chung, V. Q. (2025). Designing an AI-supported formative assessment model for pre-service mathematics teacher self-study in Vietnam. *International Journal of Interactive Mobile Technologies, 19*(22), 50–68. https://doi.org/10.3991/ijim.v19i22.57723

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*, 523–538. https://doi.org/10.1007/s11192-009-0146-3

Walkington, C. (2025). The implications of generative artificial intelligence for mathematics education. *School Science and Mathematics*, 1–10. https://doi.org/10.1111/ssm.18356

Wang, K. D., Wu, Z., Tufts, L., Wieman, C., Salehi, S., & Haber, N. (2025). Scaffold or crutch? Examining college students' use and views of generative AI tools for STEM education. *2025 IEEE Global Engineering Education Conference (EDUCON)*, 1–10. https://doi.org/10.1109/EDUCON62633.2025.11016406

Wardat, Y., Tashtoush, M. A., AlAli, R., & Saleh, S. (2024). Artificial intelligence in education: Mathematics teachers' perspectives, practices and challenges. *Iraqi Journal for Computer Science and Mathematics, 5*(1), 60–77. https://doi.org/10.52866/ijcsm.2024.05.01.004

Xing, W., Kim, T., Song, Y., Li, H., Li, C., & Kim, J. (2026). Unveiling interaction patterns between students and generative AI teachable agent: Focusing on students' agency and AI agents' authority. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.70038

Xing, W., Song, Y., Li, C., Liu, Z., Zhu, W., & Oh, H. (2025). Development of a generative AI-powered teachable agent for middle school mathematics learning: A design-based research study. *British Journal of Educational Technology, 56*, 2043–2077. https://doi.org/10.1111/bjet.13586

Yoon, H., Hwang, J., Lee, K., Roh, K. H., & 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

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, D., Wijaya, T. T., Wang, Y., Su, M., Li, X., et al. (2025). Exploring the relationship between AI literacy, AI trust, AI dependency, and 21st century skills in preservice mathematics teachers. *Scientific Reports, 15*, 14281. https://doi.org/10.1038/s41598-025-99127-0

Zhuang, Y., & Zhang, S. (2025). Integrating ChatGPT in mathematics teacher education: AI-based simulation role-playing to support practice-based teaching. *International Journal of Artificial Intelligence in Education, 35*(6), 3873–3895. https://doi.org/10.1007/s40593-025-00519-0