ISEJ Indonesian Science Education Journal

ISSN Online: 2716-3350
https://siducat.org/index.php/isej/
Published by Yayasan Darussalam Bengkulu
This article licensed under CC BY 4.0 International

OPEN ACCESS ARTICLE

Generative AI in University Mathematics Education Using the Theory–Context–Methodology Framework: A Systematic Literature Review

Ida Nuraida[1], Nurjanah[2], and Rahayu Kariadinata[3]
Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia[1,2,4]
Mathematics Education, Universitas Islan Negeri Sunan Gunung Djati, Bandung, Indonesia[3]
E-mail Corresponding: nurjanah@upi.edu

Received: June 9, 2026	Revised: August 23, 2026	Accepted: September 17, 2026

Abstract
Advances in generative AI are opening new opportunities for mathematics learning in higher education, but systematic mapping of the Theory–Context–Methodology (TCM) framework to integrate generative AI into mathematics learning remains limited. This study aims to identify the theories or conceptual frameworks, application contexts, and research methods used in studies of generative AI in higher education mathematics education. The method used is a PRISMA-based systematic literature review. The search was conducted on Scopus, Lens, and Dimensions, yielding 322 initial articles; after deduplication, screening, and eligibility assessment, 31 articles were analyzed. Article quality was assessed using the MMAT, while data were analyzed through TCM-based coding. The results indicate that generative AI is conceptualized as a pedagogical tool, a technology adoption, pedagogical competence, mathematical activity, and a cognitive partner. Its application is dominant in the classroom, general mathematics, ChatGPT, and for conceptual understanding. Methodologically, the studies were dominated by qualitative and quantitative approaches, while mixed methods and R&D/development research remained limited. These findings underscore the need for AI-based instructional designs that are critical, ethical, and supported by more integrative empirical research.

Keywords: Generative AI; mathematics education; higher education; systematic literature review; Theory–Context–Methodology.

INTRODUCTION

Mathematics learning in higher education is characterized by complexity, encompassing conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition (El-Shara et al., 2025; May, 2024). This complexity indicates that learning is not only focused on procedures but also demands higher-order thinking skills, such as reasoning, modeling, representation, and mathematical communication (Büchele & Feudel, 2023). Nevertheless, various studies indicate that students still struggle to understand abstract concepts and integrate mathematical representations deeply, leading to learning that tends to be procedure-oriented rather than focused on developing conceptual understanding and higher-order thinking (Sofroniou et al., 2025). As a result, there is a gap between the ideal goals of mathematics education and its implementation in practice.In response to these challenges, various technology-based learning innovations have been developed to improve the quality of mathematics education at the university level. Mathematics software, such as GeoGebra and MATLAB, interactive visualizations, multimedia e-textbooks, mobile learning applications, and flipped classrooms supported by interactive videos are among the most widely implemented forms of innovation (Cirneanu & Moldoveanu, 2024; Engelbrecht & Borba, 2024; St Omer et al., 2025; Sofroniou et al., 2025). Gamification and virtual simulations have also been used to enhance students' motivation and engagement in mathematics learning (Engelbrecht & Borba,

2024; Güler et al., 2022). Furthermore, the use of mathematics software, interactive visualization, and flipped classroom approaches has been shown to improve student engagement and learning outcomes (Castillo et al., 2025). In addition, global trends in mathematics education emphasize the importance of developing students' problem-solving skills and integrating technology meaningfully into the learning process (Hendriani et al., 2025).

University mathematics is a distinct GenAI context because it requires more than the production of a plausible answer. Students must coordinate formal definitions, symbolic representations, multistep deductions, modeling assumptions, and proof validation. Consequently, fluent but mathematically invalid AI output including hidden assumptions, circular arguments, unjustified implications, and symbolic errors can directly affect what is learned and assessed. A domain-specific synthesis is therefore needed to determine not only whether GenAI is used, but also which mathematical activities it mediates and how learners verify its outputs. Empirical work in undergraduate proof contexts confirms that students may use GenAI as a source of conjectures and arguments, but must critically evaluate inaccurate or unjustified mathematical claims (Yoon et al., 2024).

A number of studies indicate that GenAI has been used in various contexts of mathematics learning in higher education. Some studies position AI as a pedagogical tool to support conceptual understanding, practice, visualization, and scaffolding (Etcuban, 2025; Mairing et al., 2024; Maknun et al., 2026; Mustapa & Masaniku, 2025). Other studies examine students' acceptance of GenAI through factors such as motivation, self-efficacy, digital literacy, and perceptions of the technology's usefulness (Y. Chen et al., 2025; Davor et al., 2026). Furthermore, GenAI is also beginning to be studied as a support for prospective teachers' pedagogical competencies, a mediator of mathematical activities, and a cognitive partner in mathematical reasoning, representation, argumentation, and proof (Cevikbas & Kaiser, 2026; W.-Y. Chen, 2025; Mokgwathi & Ngoveni, 2026; Urhan et al., 2026; Yoon et al., 2024). This diversity indicates that GenAI is no longer merely a technical tool but also a pedagogical and cognitive resource in mathematics learning.

In this review, Generative AI (GenAI) is operationally defined as a computational system capable of producing new content—such as natural-language explanations, mathematical solutions, proofs, examples, representations, or feedback—in response to user prompts. Included technologies therefore comprised explicitly identified generative models, large language models (LLMs), ChatGPT, and chatbots with documented generative functions. Predictive analytics, machine-learning classifiers, conventional rule-based chatbots, automated grading systems, and intelligent tutoring systems without a generative function were excluded. This distinction is maintained throughout the manuscript; 'AI' is used only when referring to the broader field or to the terminology used by a cited source.

This operational separation between generative and non-generative AI is consistent with research distinguishing conversational content generation from predictive, classificatory, and rule-based educational AI applications (Tlili et al., 2023; Zawacki-Richter et al., 2019).

Prior reviews were identified through searches combining "generative AI," "ChatGPT," "AI chatbot," "higher education," and "systematic review." Reviews were compared by educational level, disciplinary scope, AI type, synthesis framework, and attention to mathematics-specific activity. Existing syntheses predominantly addressed cross-disciplinary higher education, personalized support, adoption, ethics, and policy; they did not jointly map theoretical assumptions, mathematical activities, and methodological evidence in university mathematics.

Recent SLR studies have also begun to highlight pedagogical aspects, user behavior, and personalized learning, but they have yet to focus specifically on university mathematics education. Dikilitaş et al. (2024), in their review of ChatGPT use by students in higher education, identified benefits for skill development, immediate feedback, motivation, and engagement, while also noting related ethical issues. Wu et al. (2025), in their review of responses, attitudes, and usage behaviors regarding GenAI in higher education teaching and learning, focused on students' and faculty's psychological responses to GenAI. Sharma et al. (2025), in their study of

LLMs in personalized learning, demonstrated the potential of LLMs to support adaptive learning, engagement, and monitoring of learning progress. Meanwhile, Qian (2025), in a review of pedagogical applications of GenAI in higher education, linked GenAI to creativity, critical thinking, learning autonomy, prompt literacy, and student–instructor–AI collaboration.

Previous reviews have largely examined AI chatbots or GenAI across disciplines in higher education, emphasizing general benefits, adoption, personalization, academic integrity, and institutional policy. They have not jointly explained (a) how GenAI is theoretically positioned in university mathematics, (b) the mathematical topics and activities in which it is used, and (c) how methodological choices shape the reported evidence. This review addresses that knowledge gap through the TCM framework, which links theoretical positioning, application context, and research methodology in a single synthesis. Accordingly, the review addresses three questions: (RQ1) What theories or conceptual frameworks underpin the use of GenAI in university mathematics learning? (RQ2) In what contexts is GenAI applied in university mathematics learning? and (RQ3) What methodological approaches characterize empirical research on GenAI in university mathematics education?

This disciplinary gap is consistent with earlier higher-education AI reviews that identified weak pedagogical theorization, limited educator involvement, and insufficient critical attention to ethical risks (Zawacki-Richter et al., 2019), as well as broader GenAI research emphasizing both pedagogical affordances and response-quality concerns (Tlili et al., 2023).

METHODS

The review protocol was developed before screening but was not prospectively registered in a public repository or independently audited; this is acknowledged as a limitation. PRISMA 2020 was used as a reporting guideline rather than as a claim that registration alone established methodological rigor. Searches were conducted in Scopus, Dimensions, and Lens on April 11, 2026, for publications from 2021 to 2026. The search operationalized Population, Intervention/phenomenon of Interest, and Outcome (PIO): higher education; GenAI; and mathematics learning or mathematical activity. A Comparison component was omitted because the review included qualitative, observational, acceptance, and design-based studies as well as interventions and did not require a common comparator. Broad AI terms were retained to maximize sensitivity because some studies used general AI terminology for generative systems; non-generative studies were removed through the operational definition during screening.

The identification, screening, retrieval, eligibility, and reporting procedures were aligned with the PRISMA 2020 statement and its emphasis on transparent reporting of information sources, selection decisions, and exclusion reasons (Page et al., 2021).

Table 1. The results of the initial search and the first stage of screening

Search terms	Database	Hits
(“higher education” OR “university” OR “tertiary education” OR “college” OR “undergraduate” OR “prospective teacher”) **AND** (“generative AI” OR “artificial intelligence” OR “large language model*” OR “LLM*” OR “ChatGPT” OR “AI chatbot*” OR “AI tutor*” OR “AI-assisted learning”) **AND** (“mathematics education” OR “mathematics learning” OR “teaching mathematics” OR “mathematical reasoning” OR “mathematical problem solving” OR “computational thinking mathematic*”)	Scopus	68
	Lens	83
	Dimensions	171
Total records before duplicate removal		322

Table 1 presents the database-specific search strategy and the number of records retrieved

before deduplication. The search terms operationalized the higher-education population, GenAI as the phenomenon of interest, and mathematics-learning outcomes. Database syntax was adapted where necessary while preserving the same conceptual blocks.

A comparison element was not required because the review synthesized heterogeneous empirical designs rather than estimating a common comparative intervention effect. The three PIO blocks were combined with AND, while synonyms within each block were combined with OR. Broad AI-related terms improved sensitivity; studies were retained only when full-text evidence met the operational GenAI definition.

Table 2. Inclusion and Exclusion Criteria

Category	Inclusion Criteria	Exclusion Criteria
Year	Published between 2021 and 2026	Published before 2021
Topic	Direct use or evaluation of GenAI in university mathematics teaching, learning, reasoning, problem solving, modeling, representation, assessment, or proof	Non-generative AI; no direct GenAI use; or no explicit mathematics teaching/learning activity
Subject	Higher-education students, prospective teachers, or university mathematics instructors	Participants exclusively outside higher education or without a university mathematics context
Type of Study	Primary empirical research collecting and analyzing participant, performance, interaction, assessment, observation, interview, survey, or AI-output data	Review, meta-analysis, conceptual paper, opinion, editorial, protocol without results, or non-empirical commentary
Quality	Peer-reviewed articles with full text available	Non-peer-reviewed articles or full text unavailable
Language	English	Languages other than English

An empirical study was defined as primary research that systematically collected and analyzed data from participants, learning activities, assessments, interviews, observations, surveys, AI-generated outputs, or interaction logs. Mathematics-education relevance required an explicit higher-education setting, a mathematics teaching or learning activity, and direct use or evaluation of GenAI by students, prospective teachers, or instructors. Screening was conducted sequentially at title, abstract, and full-text levels. At title–abstract screening, records had to indicate higher education, mathematics, and an AI application potentially meeting the GenAI definition. At full-text screening, the reviewers verified the generative function, participant context, empirical design, mathematics-learning relevance, publication type, language, and full-text availability. Each excluded report was assigned one primary reason to prevent double counting.

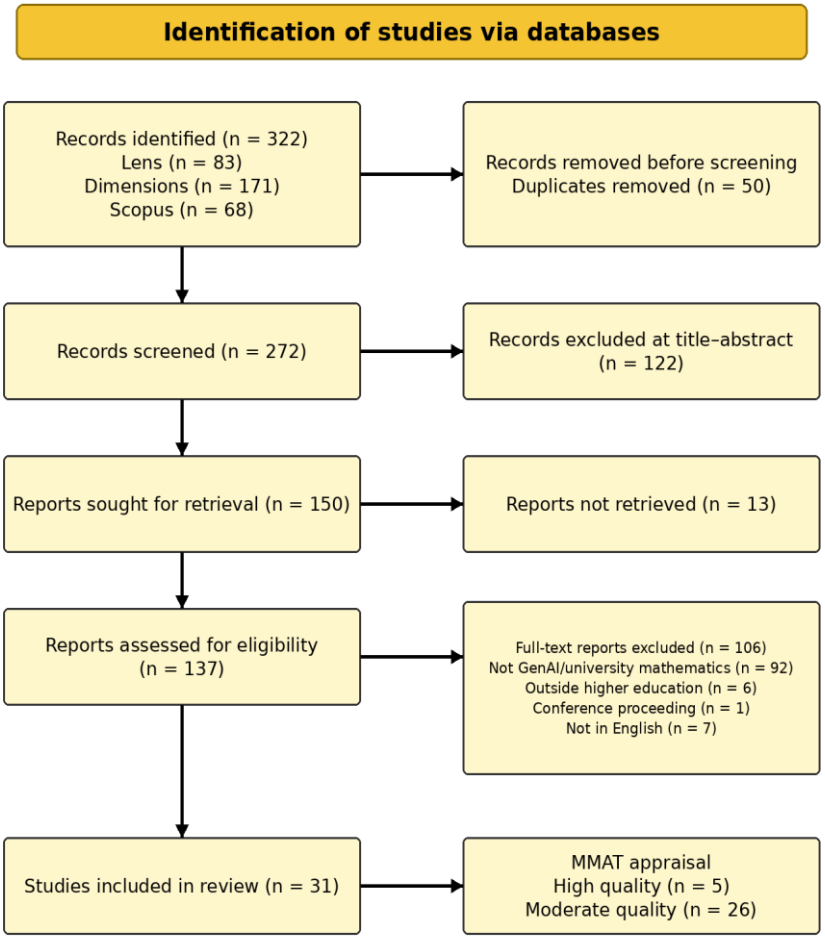


Figure 1. PRISMA Flow Chart Of Article Identification And Screening

The searches identified 322 records (Lens, n = 83; Dimensions, n = 171; Scopus, n = 68). After 50 duplicates were removed, 272 records were screened and 122 were excluded at title–abstract screening. Of 150 reports sought for retrieval, 13 could not be obtained; 137 full-text reports were assessed for eligibility. A further 106 reports were excluded: 92 did not meet the operational criteria for GenAI use and/or university mathematics education, 6 involved participants outside higher education, 1 was a conference proceeding, and 7 were not written in English. The remaining 31 studies were included in the synthesis. Full-text exclusions were assigned one primary reason and itemized in the review records.

Methodological quality was appraised using the MMAT design-specific criteria. Two reviewers assessed the included studies independently and resolved disagreements through discussion. Because a numerical inter-reviewer agreement coefficient was not recorded during the original appraisal, no kappa value is claimed; this reduces the reproducibility of the appraisal and is reported as a limitation. MMAT findings were used to qualify interpretation rather than to exclude otherwise eligible studies. The appraisal identified 5 studies with consistently strong evidence across the applicable criteria and 26 with one or more limitations; no study was excluded solely on quality grounds. Recurring concerns included context-specific or convenience samples, short observation periods, reliance on self-reported perceptions, limited validation of outcome measures, and insufficient integration of qualitative and quantitative components.

The MMAT appraisal yielded 5 studies with consistently strong evidence across the applicable design criteria and 26 studies with one or more methodological limitations. Because appraisal informed interpretation rather than eligibility, all 31 studies remained in the synthesis. Findings from moderate-quality studies were interpreted cautiously, particularly where conclusions relied on self-report, small or convenience samples, short interventions, or

incompletely validated measures.

The individual empirical study was the unit of coding. Evidence was extracted from the theory, methods, results, and discussion sections into a coding manual containing an operational definition, inclusion and exclusion rules, and an example for every TCM category. For example, a study was coded T1 when GenAI provided explanations, practice, feedback, visualization, or scaffolding; T5 required evidence that learners used GenAI in reasoning, argumentation, proof, reflection, or critical evaluation. Multiple coding was allowed for non-mutually exclusive variables, particularly purpose of use (C4), because one study could address conceptual understanding, engagement, and assessment simultaneously. Methodological design was assigned one primary category, with integrated qualitative and quantitative designs coded as mixed methods. Two reviewers coded the studies independently and resolved discrepancies through consensus. A numerical intercoder-reliability coefficient was not retained in the original coding records and is therefore not reported; the study-level coding matrix is provided to make category assignments auditable.

Coding examples and decision rules were added to the accompanying methodological explanation, while the study-level matrix reports how every included article was assigned to the TCM categories. Multiple codes are separated by semicolons where a study addressed more than one purpose of GenAI use.

Table 3. Variables and Codes Used in the Study

Code	**Dimension**	**Description**
T1	Pedagogical Tool	AI as a learning support tool
T2	Technology Adoption	AI as an adopted technology
T3	Pedagogical Competence	AI as a means of supporting pedagogical competence
T4	Mathematical Activity	AI as a mediator of mathematical activities
T5	Cognitive Partner	AI as a cognitive partner
C1	Learning Environment	C1.1 Classroom; C1.2 Blended Learning; C1.3 Online Learning; C1.4 Laboratory-Based Learning
C2	Mathematics Topic	C2.1 General Mathematics; C2.2 Algebra; C2.3 Calculus; C2.4 Geometry; C2.5 Trigonometry; C2.6 Real Analysis; C2.7 Mathematical Modeling
C3	Type of Generative AI	C3.1 ChatGPT; C3.2 Large Language Models; C3.3 AI Chatbots; C3.4 Combination of AI Tools
C4	Purpose of AI Use	C4.1 Problem Solving; C4.2 Mathematical Proof; C4.3 Conceptual Understanding; C4.4 Assessment; C4.5 Learner Engagement; C4.6 AI Literacy
M1	Qualitative Study	Research that explores experiences, perceptions, meanings, or processes in depth using non-numerical data, such as interviews, observations, or documents.
M2	Quantitative Study	Research that measures variables, relationships, or effects using numerical data and statistical analysis.
M3	Mixed Methods Study	Research that combines qualitative and quantitative data to obtain a more comprehensive understanding.
M4	R&D / Development Research	Designs, develops, validates, and tests educational products, models, or AI-based learning tools.

RESULTS AND DISCUSSION

Result

The analysis findings are organized to provide a systematic overview of trends in generative AI research in mathematics education at the higher education level. The analysis was conducted across three main dimensions: Theory, Context, and Methodology. The Theory dimension examines the theoretical foundations underpinning the research; the Context dimension maps the learning environment, mathematical topics, types of generative AI, and the objectives of their

use; while the Methodology dimension identifies the research designs employed. Through this coding, patterns of research development can be seen more clearly, from conceptual, pedagogical, and methodological perspectives. A more comprehensive overview can be found in Table 4.

Beyond frequency patterns, the synthesis indicates that reported benefits depended on pedagogical and learner-level mechanisms. GenAI supported immediate explanation, example generation, feedback, and dialogic scaffolding, but these functions were mediated by prior mathematical knowledge, prompt quality, and students' capacity to verify generated output. Studies framing GenAI as a cognitive partner more often foregrounded reasoning, representation, argumentation, and critique, whereas technology-adoption studies primarily measured intention, perceived usefulness, or self-efficacy. Thus, positive perceptions should not be interpreted as direct evidence of improved mathematical competence. Across categories, short interventions, single-institution samples, and self-report data limited causal and long-term conclusions.

Table 4. Coding Framework for Extracted Studies Based on Theory, Context, and Methodology

No	Code	Author	Theory	Context				Method
				C1	C2	C3	C4	
1	A1	(Maknun et al., 2026)	T1	C1.1	C2.5	C3.1	C4.3; C4.5	M4
2	A2	(Mokgwathi & Ngoveni, 2026)	T3	C1.3	C2.1	C3.2	C4.4; C4.9	M3
3	A3	(Davor et al., 2026)	T2	C1.1	C2.1	C3.1	C4.1; C4.9	M2
4	A4	(Urhan et al., 2026)	T5	C1.1	C2.3	C3.1	C4.1; C4.3	M1
5	A5	(Spring et al., 2026)	-	C1.1	C2.7	C3.2	C4.1; C4.7; C4.9	M3
6	A6	(Pala et al., 2025)	-	C1.1	C2.2	C3.4	C4.2	M3
7	A7	(Lau et al., 2025)	-	C1.1	C2.1	C3.1	C4.8; C4.9	M2
8	A8	(W. Lin & Jiang, 2025)	T2	C1.1	C2.1	C3.1	C4.4; C4.3	M2
9	A9	(Li, 2025)	T2	C1.1	C2.1	C3.3	C4.1; C4.3; C4.5	M2
10	A10	(Divjak et al., 2026)	-	C1.1	C2.1	C3.4	C4.1; C4.5	M2
11	A11	(Brandibur et al., 2026)	-	C1.1	C2.1	C3.2	C4.5; C4.9	M2
12	A12	(Zhuang, 2025)	T3	C1.1	C2.1	C3.3	C4.5	M1
13	A13	(Hasratuddin et al., 2026)	T1	C1.1	C2.1	C3.1	C4.1; C4.4	M4
14	A14	(Mustapa & Masaniku, 2025)	T1	C1.1	C2.1	C3.4	C4.3; C4.9	M1
15	A15	(Ergene & Caylan-Ergene, 2026)	-	C1.1	C2.7	C3.1	C4.1; C4.7	M2
16	A16	(Sucipto & Negara, 2023)	-	C1.4	C2.1	C3.2	C4.3; C4.7	M2
17	A17	(Mairing et al., 2024)	T1	C1.1	C2.3	C3.1	C4.3	M1
18	A18	(Matzakos et al., 2023)	-	C1.1	C2.4	C3.3	C4.3	M3
19	A19	(Spring et al., 2026)	-	C1.1	C2.1	C3.4	C4.3	M1
20	A20	(Yu et al., 2026)	T3	C1.1	C2.1	C3.1	C4.5; C4.9	M3
21	A21	(Etcuban, 2025)	T1	C1.1	C2.3	C3.1	C4.3	M2
22	A22	(Vankúš, 2024)	-	C1.1	C2.2	C3.3	C4.8	M2
23	A23	(Serhan & Welcome, 2024)	-	C1.1	C2.1	C3.1	C4.1; C4.6	M3
24	A24	(Moral-Sánchez et al., 2023)	-	C1.1	C2.2	C3.1	C4.5	M2
25	A25	(Cevikbas & Kaiser, 2026)	T4	C1.1	C2.1	C3.1	C4.4	M1
26	A26	(Lagon, 2025)	T5	C1.2	C2.6	C3.1	C4.2	M1
27	A27	(Matzakos & Moundridou, 2025)	-	C1.1	C2.1	C3.1	C4.2; C4.9	M1
28	A28	(Urhan et al., 2025)	T4	C1.1	C2.1	C3.1	C4.9	M1
29	A29	(Yoon et al., 2024)	T5	C1.1	C2.3	C3.1	C4.3; C4.5	M1
30	A30	(Supriyadi et al., 2025)	-	C1.1	C2.1	C3.2	C4.1; C4.3	M1
31	A31	(Filiz & Gür, 2025)	-	C1.2	C2.4	C3.3	C4.5; C4.9	M3

Theoretical and Conceptual Foundations of Generative AI in Mathematics Learning

The results of the analysis indicate that the conceptualization of generative AI in mathematics education at the university level can be grouped into five main dimensions: as a pedagogical tool, adopted technology, pedagogical competencies, a mediator of mathematical activities, and a cognitive partner. In general, the distribution of studies indicates that the Pedagogical Tool and Cognitive Partner dimensions are the most dominant, each represented by 4 studies, followed by the other dimensions, with a relatively balanced distribution.

Table 5. Conceptual Framework

Code	Theory Dimension	Theory / Conceptual Framework	Article Codes
T1	Pedagogical Tool	Constructivism, multimedia learning, problem-based learning, inquiry-based learning, cognitive load theory, self-efficacy, ZPD	A1, A14, A17, A21
T2	Technology Adoption	TAM, UTAUT, social cognitive theory, motivation theory	A3, A8, A9
T3	Pedagogical Competence	AI literacy, GenAI literacy, TPACK, formative assessment	A2, A12, A20
T4	Mathematical Activity	Realistic mathematics education, modeling theory, semiotic representation theory	A13, A25, A28
T5	Cognitive Partner	Toulmin's argumentation model, rationality, instrumental approach, SIPE-AI, metacognition	A4, A15, A26, A29

The results of the data extraction indicate that, of the 31 articles analyzed, 17 explicitly used a theory or conceptual framework to explain the use of generative AI in mathematics education. These frameworks are divided into five main dimensions. The Pedagogical Tool dimension (T1) appears in A1, A14, A17, and A21, positioning GenAI as a pedagogical tool to support concept explanation, practice, visualization, feedback, and scaffolding. The Technology Adoption dimension (T2) is evident in A3, A8, and A9, highlighting factors such as perceived usefulness, ease of use, motivation, digital literacy, self-efficacy, and students' readiness to utilize AI for learning mathematics. Meanwhile, the Pedagogical Competence dimension (T3) is evident in A2, A12, and A20, linking GenAI to the development of pedagogical competencies, AI literacy, formative assessment, lesson planning, and prospective teachers' readiness to use AI responsibly.

Furthermore, the Mathematical Activity dimension (T4) is found in A13, A25, and A28, which position GenAI as a mediator in mathematical activities, such as representation, modeling, investigation, and mathematical reasoning. The Cognitive Partner dimension (T5) appears in A4, A4, A15, A26, and A29, conceptualizing GenAI as a cognitive partner in dialogue, argumentation, reflection, proof, and critical evaluation of AI responses. Overall, these results indicate that the use of GenAI in mathematics learning is not merely understood as a technical tool but has evolved into a pedagogical resource, an adopted technology, a supporter of pedagogical competencies, a mediator of mathematical activities, and a cognitive partner in students' thinking processes.

Mapping the Contexts of Generative AI Application in Higher Education Mathematics Learning

The results of the data extraction for RQ2 indicate that the context of generative AI implementation in mathematics education in higher education can be categorized into four main categories: Learning Environment (C1), Mathematics Topic (C2), Type of Generative AI (C3), and Purpose of AI Use (C4). In category C1, the majority of articles fall under C1.1 Classroom, namely 27 articles, while C1.2 Blended Learning appears in 2 articles, C1.3 Online Learning in 1 article, and C1.4 Laboratory-based Learning in 1 article (Figure 2).

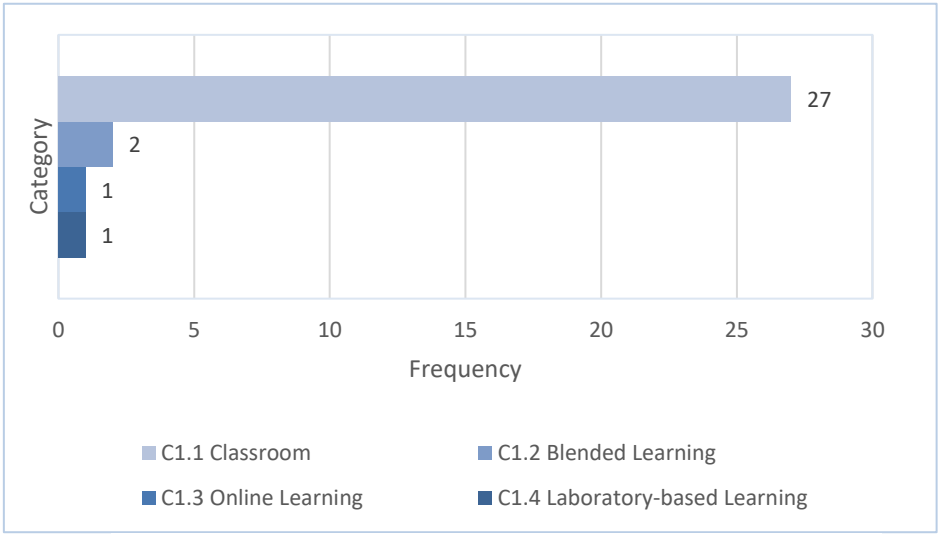


Figure 2. Learning Environment (C1)

Figure 2 shows the distribution of the Learning Environment (C1) category in research on generative AI in mathematics education at the higher education level. The most dominant category is C1.1 Classroom, with 27 articles, indicating that the majority of studies apply or examine generative AI in classroom learning contexts. Meanwhile, C1.2 Blended Learning appears in only 2 articles, while C1.3 Online Learning and C1.4 Laboratory-based Learning each appear in 1 article. These findings show that the use of generative AI is still more integrated into conventional classroom learning, while its application in blended, fully online, and mathematics laboratory learning remains relatively limited.

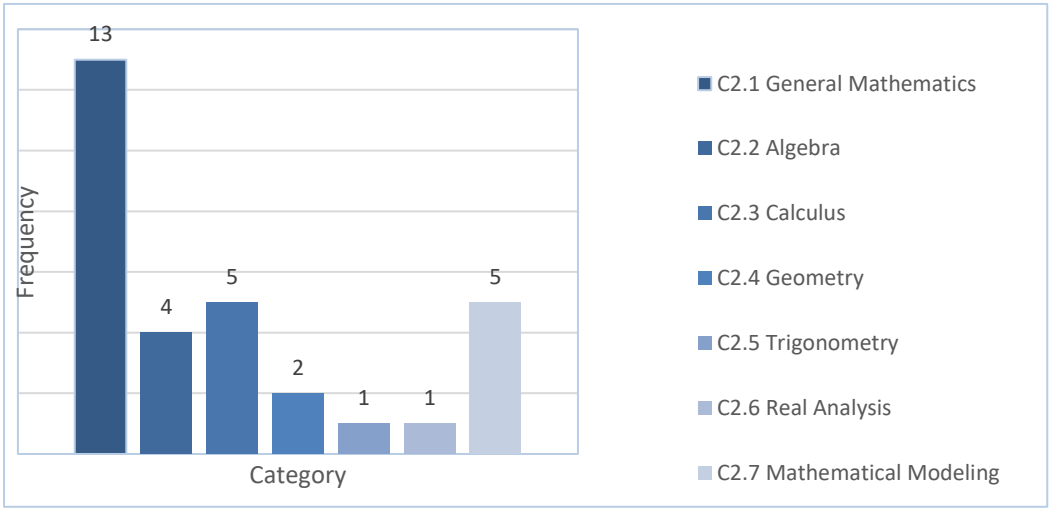


Figure 3. Mathematics Topic (C2)

Figure 3 shows the distribution of Mathematics Topics (C2) in the analyzed articles. The most dominant category is C2.1 General Mathematics, with 13 articles, indicating that most of the research discusses the use of generative AI in mathematics education broadlybroadly, without limiting itself to a specific topic. Next, C2.3 Calculus and C2.7 Mathematical Modeling each appear in 5 articles, followed by C2.2 Algebra with 4 articles. Meanwhile, C2.4 Geometry appears in 2 articles, while C2.5 Trigonometry and C2.6 Real Analysis each appear in only 1 article. Overall, this diagram indicates that GenAI research remains dominant in general mathematics but is beginning to expand into specific areas such as calculus, algebra, geometry, real analysis, and mathematical modeling.

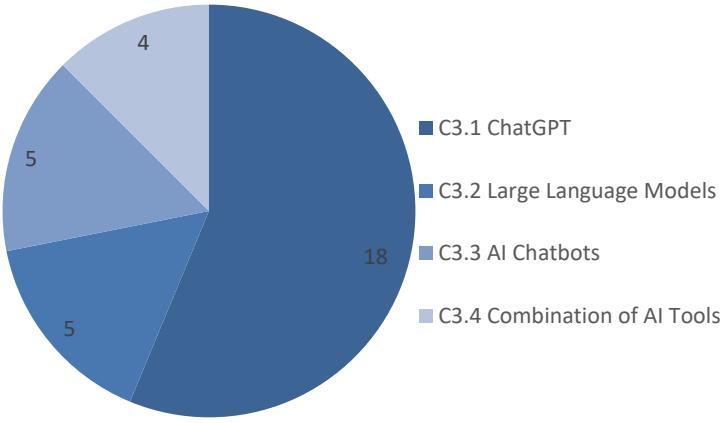


Figure 4. Type of Generative AI (C3)

Figure 4 shows the distribution of the Type of Generative AI (C3) used in the analyzed articles. The most dominant category is C3.1 ChatGPT with 18 articles, making ChatGPT the most frequently studied generative AI platform in mathematics education at the higher education level. Next, C3.2 Large Language Models and C3.3 AI Chatbots each appeared in 5 articles, while C3.4 Combination of AI Tools appeared in 4 articles. These results indicate that research remains heavily focused on ChatGPT but is beginning to expand into broader studies, whether through LLMs in general, AI chatbots, or combinations of multiple AI tools such as ChatGPT, Photomath, Gemini, Copilot, and others.

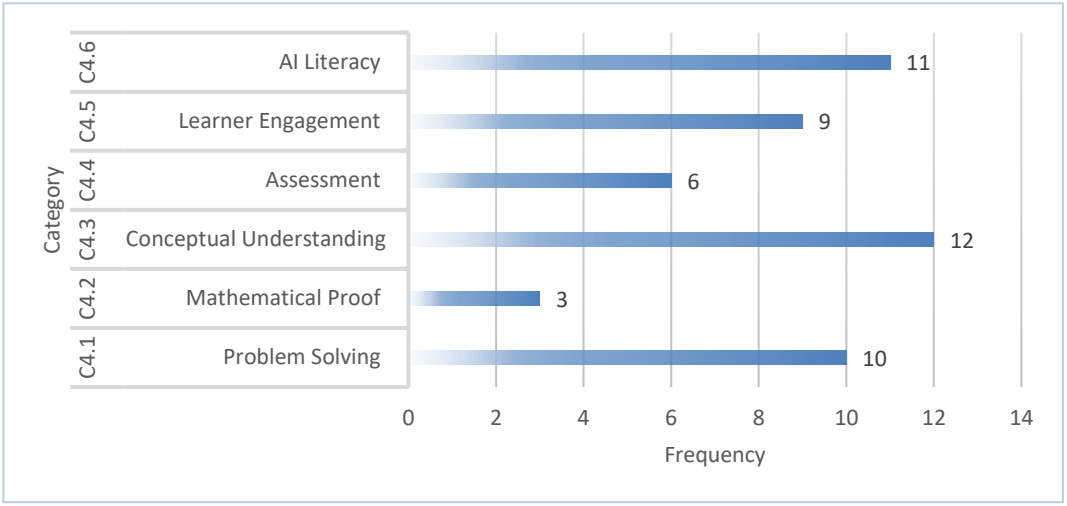


Figure 5. Purpose of AI Use (C4)

Figure 5 shows the distribution of Purpose of AI Use (C4) in the analyzed articles. The most dominant category is C4.3 Conceptual Understanding with 12 articles, followed by C4.6 AI Literacy with 11 articles, and C4.1 Problem Solving with 10 articles. Category C4.5 Learner Engagement appears in 9 articles, indicating that AI is also frequently associated with enhancing students' motivation, interest, self-confidence, and engagement in learning mathematics. Meanwhile, C4.4 Assessment appeared in 6 articles, whereas C4.2 Mathematical Proof appeared in only 3. Thus, the use of AI in mathematics education remains more prevalent in conceptual understanding, AI literacy, and problem-solving, while its application to mathematical proof and assessment remains relatively limited.

Methodological Trends in Studies on Generative AI in Higher Education Mathematics Learning

Methodological trends in the study "Generative AI in University Mathematics Education Using the Theory–Context–Methodology Framework: A Systematic Literature Review" were analyzed to understand how previous studies designed, collected, and interpreted empirical evidence regarding the use of generative AI in mathematics learning in higher education. Through the Methodology dimension of the TCM framework, this study maps trends in research approaches, including qualitative, quantitative, mixed-methods, and R&D/development research. This analysis is important because the choice of method not only indicates how researchers measure effectiveness or explore user experiences, but also reflects the maturity of the research in explaining the processes, impacts, and challenges of integrating generative AI pedagogically.

Table 6. Methodological Category

Method Category	Article Codes	Number
Qualitative Study	A4, A12, A14, A18, A19, A25, A26, A27, A28, A29, A30	11
Quantitative Study	A3, A7, A8, A9, A10, A11, A15, A16, A21, A22, A24	11
Mixed Methods Study	A2, A5, A6, A20, A23, A31	6
R&D / Development Research	A1, A13, A17	3
Total		**31**

Table 6 shows the distribution of research methods among the 31 analyzed articles. Quantitatively, the Qualitative Study and Quantitative Study categories were the most dominant methods, each comprising 11 articles. Qualitative studies included articles A4, A12, A14, A18, A19, A25, A26, A27, A28, A29, and A30, while quantitative studies include A3, A7, A8, A9, A10, A11, A15, A16, A21, A22, and A24. Meanwhile, Mixed Methods Studies were found in 6 articles: A2, A5, A6, A20, A23, and A31. The R&D/Development Research category has the fewest number of articles, namely 3, consisting of A1, A13, and A17. Thus, research on generative AI in mathematics learning in higher education is dominated by qualitative and quantitative approaches, while mixed methods and R&D/development research remain relatively limited.

The methodological pattern also reveals differences in evidential strength that are obscured by frequency counts alone. Qualitative studies illuminated prompting practices, trust, reasoning, and verification but were typically based on small, context-specific samples. Quantitative studies more often examined acceptance, perceptions, confidence, or performance, yet many relied on convenience samples and self-report measures without longitudinal follow-up or a comparison condition. Mixed-methods studies offered broader evidence but did not always demonstrate explicit integration between qualitative and quantitative findings. Development studies produced instructional tools or designs, although evaluation was commonly preliminary. These limitations indicate risks of selection bias, common-method bias, and overgeneralization from short-term outcomes.

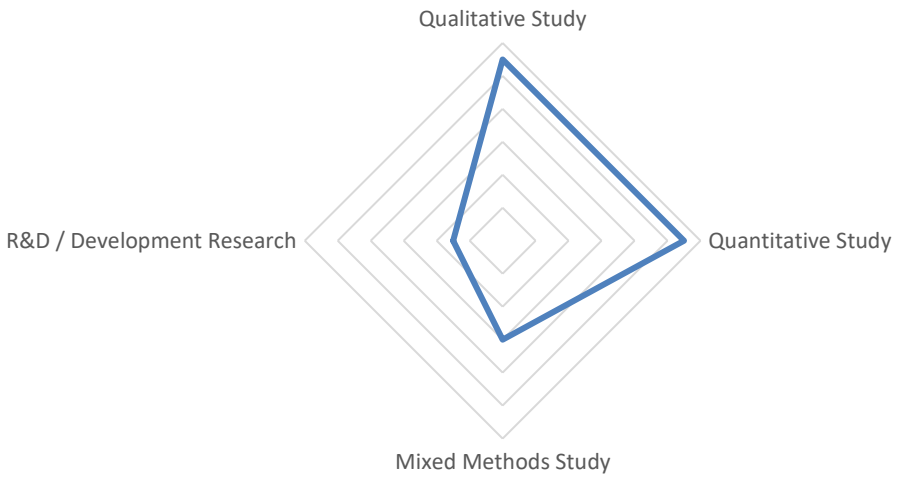


Figure 6. Trends Methodology

Figure 6 shows that the methodological trends in research on generative AI in mathematics education at the higher education level are dominated by qualitative and quantitative approaches, which are visually the most prominent and relatively balanced. This indicates that GenAI studies focus not only on exploring students' experiences, perceptions, interactions, and learning processes, but also on empirically measuring perceptions, technology acceptance, learning outcomes, and the effectiveness of AI use. Meanwhile, Mixed Methods Studies are at an intermediate level, indicating efforts to combine qualitative and quantitative data to gain a more comprehensive understanding. As for R&D/Development Research, it has the lowest prevalence, suggesting that research focused on the development, validation, and testing of GenAI-based learning models or tools remains relatively limited and presents a significant opportunity for future research.

Discussion

This review synthesized 31 empirical studies to examine how generative artificial intelligence (GenAI) is conceptualized, applied, and investigated in university mathematics education. The findings identify five conceptual orientations: pedagogical support, technology adoption, pedagogical competence, mathematical activity, and cognitive partnership. Applications concentrate on classroom settings, general mathematics, and ChatGPT, while qualitative and quantitative approaches predominate over mixed-methods and development research. Across these dimensions, the central interpretive distinction concerns assistance with completing mathematical tasks versus evidence of independent mathematical understanding. Thus, the review addresses its three research questions by connecting theoretical positioning with the activities examined and the methodological strength of the claims made, rather than treating publication frequencies as evidence of effectiveness.

The concentration on accessible instructional support accords with Lo's (2023) rapid review, which identified opportunities for feedback and resource generation alongside variable accuracy across disciplines. However, the present synthesis narrows the evaluative criterion to mathematical explanation, representation, reasoning, and proof. Crompton and Burke (2023) mapped a broader higher-education AI field that included prediction, assessment, and intelligent tutoring; those applications cannot automatically be treated as GenAI evidence. This distinction clarifies why a narrower corpus may reveal different application patterns. Moreover, Bond et al. (2024) called for stronger ethical, contextual, and methodological grounding in higher-education AI research. The predominance of exploratory and perception-oriented evidence in the present review reinforces that concern without establishing that particular research designs are intrinsically inadequate.

The cognitive-partner orientation is particularly meaningful when learners retain responsibility for verification. Yoon et al. (2024) showed that students' conceptions of proof, conceptions of GenAI, and ethical considerations shaped their proving practices. Similarly,

Cevikbas and Kaiser (2026) found that preservice teachers valued LLM assistance during modeling but questioned its accuracy and usefulness for validation. Urhan et al. (2026) further illustrated differences between two students in detecting and using erroneous calculus responses. These studies support interpreting partnership as a contingent learning relationship rather than an inherent property of the technology. Consistent with Kasneci et al. (2023), explanatory fluency offers educational possibilities but also creates demands for subject knowledge and critical evaluation. In mathematics, therefore, perceived helpfulness and coherent language are insufficient criteria for epistemic reliability.

International outcome studies also qualify optimistic adoption narratives. Chan and Hu (2023) documented generally positive university-student perceptions, whereas Kestin et al. (2025) measured greater short-term learning gains with a deliberately structured AI tutor than with an active-learning condition in undergraduate physics. In contrast, Bastani et al. (2025) found that unrestricted AI assistance could improve supported practice while undermining subsequent unaided performance in high-school mathematics; pedagogical safeguards mitigated this risk. These studies differ in educational level, subject, tutor design, and outcome timing, so their findings are not directly contradictory. Together with Zhai et al.'s (2024) review of overreliance, they suggest that the present corpus should be interpreted through the distinction between immediate task performance and retained competence. Neither favorable perceptions nor short-term gains establish durable mathematical learning.

Context provides a further explanation for variation. Al Maqbali (2025), in a two-class observational study at an Omani university, described more teacher-centered mathematics instruction than English instruction. Although the study did not evaluate GenAI, it illustrates how disciplinary classroom practices may condition opportunities for dialogue and independent critique. Such contextual differences offer a more defensible explanation than attributing outcomes to national culture without comparative evidence. Yan et al. (2024) also identified practical and ethical weaknesses in LLM applications, while Farrokhnia et al. (2024) emphasized competing educational opportunities and threats. Accordingly, platform availability, instructional organization, and verification demands should be examined together when interpreting differences across settings.

Indonesian studies reinforce the importance of instructional design while demonstrating why findings require careful qualification. Mairing et al. (2024) developed a problem-, YouTube-, and ChatGPT-supported model for real-analysis proving and reported higher scores in the PYChat class than in the comparison class. However, the intervention combined several components, preventing attribution of the reported advantage to ChatGPT alone. Maknun et al. (2026) likewise described engaging AI-assisted trigonometry media but identified confusion about the tangent function and difficulties with self-directed learning. Mustapa and Masaniku (2025), drawing on three students in Gorontalo, reported different perceived affordances of Photomath and ChatGPT-4o. Their phenomenological evidence illuminates experience rather than establishing comparative effectiveness. Collectively, these studies align with the review's emphasis on conceptual support, but their differing tasks, instructional packages, and samples caution against a single national effectiveness claim.

Adjacent Indonesian evidence also helps distinguish confidence from demonstrated reasoning. Hidayah and Sodikin (2026), in undergraduate physics, reported higher academic-confidence and decision-making questionnaire scores under an explanatory AI interface. Nevertheless, allocation across only three course clusters, unmodeled course dependence, and self-report outcomes constrain causal and transfer claims. The study does not demonstrate that confidence is calibrated to mathematical accuracy. Its relevance is therefore to future measurement and design: university mathematics research should examine whether explanations improve justified decisions, rather than assuming that increased confidence signals improved competence.

Theoretically, the findings suggest that adoption frameworks should be connected to accounts of mathematical activity and metacognitive verification: willingness to use GenAI

explains participation, whereas learning requires evidence of how students evaluate and reconstruct mathematical arguments. TCM provides an organizing lens for this relationship, not a causal theory validated by the review. Pedagogically, educators can require independent attempts, annotated corrections of AI responses, comparison of alternative proofs, and oral justification, making engagement intellectually accountable. Supportive feedback and opportunities to discuss uncertainty may also help address anxiety, although emotional well-being was not established as a review-level outcome. Critical AI and media literacy should encompass source checking, disclosure, and responsibility for digital authorship, supporting an identity as a mathematical reasoner. At the policy level, Chan's (2023) pedagogical, governance, and operational framework and Cotton et al.'s (2024) integrity analysis support clear permitted-use rules, privacy safeguards, equitable access, and process-based assessment. Mardianti et al. (2026) further propose pedagogical, professional, social, and ethical educator competencies; this conceptual recommendation supports considering sustained professional development, rather than proving its effectiveness. Institutions should provide human support for AI-related distress and evaluate its necessity and outcomes, while avoiding mental-health or social-media regulatory claims that the reviewed mathematics evidence cannot substantiate.

The contribution is a discipline-specific synthesis that connects what GenAI is assumed to do with the mathematical activity investigated and the evidence used to support conclusions. Rather than introducing TCM as a new framework or asserting unprecedented coverage, the review applies it to distinguish adoption, instructional assistance, and cognitively substantive participation in university mathematics. Its methodological contribution lies in making theoretical orientations and study designs jointly interpretable, thereby exposing where claims about reasoning or understanding rest primarily on perception measures. This perspective refines evaluation priorities toward verification, independent performance, and transfer, and provides a basis for designing more coherent research across mathematical topics and institutional settings.

Several limitations qualify these interpretations. The review searched Scopus, Dimensions, and Lens on April 11, 2026, and included English-language, peer-reviewed journal articles, creating possible database, language, and publication biases. Rapid model changes and inconsistent terminology further limit temporal generalizability. Heterogeneous designs and outcomes precluded meta-analysis, while self-report, context-specific samples, short observation periods, and methodological weaknesses in the included studies restrict causal and cross-cultural conclusions. The protocol was not prospectively registered or independently audited, and numerical coding- and appraisal-agreement coefficients were not retained, reducing reproducibility despite consensus procedures. Future reviews should preregister protocols, retain agreement statistics, publish auditable extraction matrices, and update searches. Primary research should compare safeguarded and unrestricted use through longitudinal or appropriately controlled designs, documenting model versions, prompts, and instructional conditions. Delayed unaided assessments, proof-validation tasks, confidence calibration, and engagement and well-being measures would clarify whether support produces durable learning. Multi-institutional Indonesian and international comparisons could then examine contextual mechanisms directly.

CONCLUSION

This review was limited to Scopus, Lens, and Dimensions and to English-language, peer-reviewed journal articles; relevant conference papers, preprints, and grey literature may therefore have been omitted. Rapid changes in GenAI models and unstable terminology mean that the evidence represents a time-bounded snapshot and that some relevant studies may not have used searchable GenAI labels. Heterogeneity in participants, mathematical topics, interventions, and outcomes precluded meta-analysis and limits direct comparison. The protocol was not prospectively registered or independently audited. In addition, numerical coefficients for intercoder agreement and MMAT reviewer agreement were not retained, although coding and

appraisal disagreements were resolved through discussion. These limitations warrant cautious interpretation of frequency patterns and reported effects.

From the theoretical dimension, this study concludes that generative AI in mathematics learning in higher education is conceptualized in a multidimensional manner. GenAI is not only positioned as a technical aid but also as a pedagogical tool, a means of technology adoption, a component of pedagogical competence, a form of mathematical activity, and a cognitive partner. These five orientations indicate that GenAI can support concept explanation, feedback provision, scaffolding, mathematical representation, modeling, argumentation, reflection, and mathematical proof. Thus, in theory, GenAI has the potential to broaden how students understand mathematical concepts and engage in mathematical thinking processes.

From the Context dimension, findings indicate that the application of generative AI remains most dominant in classroom settings, general mathematics topics, the use of ChatGPT, and pedagogical objectives such as conceptual understanding, problem-solving, learner engagement, and AI literacy. This suggests that the integration of GenAI into higher mathematics education is still widely used to support regular learning, provide explanations, solve problems, and enhance learning engagement. However, its application in more specific contexts—such as algebra, calculus, geometry, mathematical modeling, online learning, blended learning, and laboratory-based learning still needs to be expanded to make GenAI usage more diverse and contextually relevant.

From the Methodology dimension, the analyzed studies indicate that qualitative and quantitative approaches are used in roughly equal proportions, while mixed-methods and R&D/development research remain more limited. This pattern suggests that GenAI research in mathematics education has begun to examine student experiences while empirically measuring the impact of AI use. However, future research needs to strengthen mixed-methods designs to combine learning outcomes data and user experiences, as well as development research to produce pedagogically and empirically validated AI-based models, modules, prompt scaffolding, assessment rubrics, and learning tools.

REFERENCE

Al Maqbali, A. (2025). Studies the dynamic of the teaching strategies in English and mathematic classrooms. Digital Learning, Social Science, and Life-course Studies, 1(2), 94–106. https://doi.org/10.70211/disolife.v1i2.339

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), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21(1), Article 4. https://doi.org/10.1186/s41239-023-00436-z

Brandibur, O., Filipowicz-Chomko, M., Girejko, E., Kaslik, E., Mozyrska, D., Mureşan, R., Pappas, N., Tănasie, A. L., & Zaharia, C. (2026). Higher mathematics education and AI prompt patterns: Examples from selected university classes. Applied Sciences, 16(1), Article 339. https://doi.org/10.3390/app16010339

Büchele, S., & Feudel, F. (2023). Changes in students' mathematical competencies at the beginning of higher education within the last decade at a German university. International Journal of Science and Mathematics Education, 21(8), 2325–2347. https://doi.org/10.1007/s10763-022-10350-x

Castillo, E. R., León, J. A. M., Celiz, E. B., Alarcón, T. J. J., & Del Milagro Chunga More, Y. (2025). Technological tools used in mathematical thinking at the university level: A systematic review. Journal of Educational and Social Research. https://doi.org/10.36941/jesr-2025-0043

Cevikbas, M., & Kaiser, G. (2026). The use of large language models to solve mathematical modeling problems: Preservice mathematics teachers' use practices, perceived affordances and challenges, and trustworthiness judgments of AI-generated outputs. ZDM – Mathematics Education, 58(5), 911–927. https://doi.org/10.1007/s11858-026-01786-4

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and

learning. International Journal of Educational Technology in Higher Education, 20(1), Article 38. https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20(1), Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chen, W.-Y. (2025). Integrating GAI tools and scaffolding-based learning to enhance learning outcomes in English-medium programming courses. Contemporary Research in Education and English Language Teaching, 7(1), 54–62. https://doi.org/10.55214/26410230.v7i1.5627

Chen, Y., Wang, Y., Wüstenberg, T., Kizilcec, R., Fan, Y., Li, Y., Lu, B., Yuan, M., Zhang, J., Zhang, Z., Geldsetzer, P., Chen, S., & Bärnighausen, T. (2025). Effects of generative artificial intelligence on cognitive effort and task performance: Study protocol for a randomized controlled experiment among college students. Trials, 26. https://doi.org/10.1186/s13063-025-08950-3

Cirneanu, A.-L., & Moldoveanu, C.-E. (2024). Use of digital technology in integrated mathematics education. Applied System Innovation, 7(4), Article 66. https://doi.org/10.3390/asi7040066

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20(1), Article 22. https://doi.org/10.1186/s41239-023-00392-8

Davor, I., Boateng, F. O., Osei-Tutu, P., Osei-Bosompem, A., Patience, A. S., & Adusei, L. A. (2026). Examining the mediating roles of digital literacy and critical thinking in the relationship between ChatGPT use and university students' mathematical problem-solving efficacy. Journal of Mathematics Science and Technology Education, 3(1), 1–15. https://doi.org/10.12973/jmste.3.1.1

Dikilitaş, K., Furenes Klippen, M. I., & Keles, S. (2024). A systematic rapid review of empirical research on students' use of ChatGPT in higher education. Nordic Journal of Systematic Reviews in Education, 2(1). https://doi.org/10.23865/njsre.v2.6227

Divjak, B., Svetec, B., & Horvat, D. (2026). Generative AI in mathematics education: Analysing student performance and perceptions over three academic years. International Journal of Technology Enhanced Learning, 18(1), 16–39. https://doi.org/10.1504/IJTEL.2026.150520

El-Shara, I. A. H., Tabieh, A. A. S., & Abu Helu, S. Y. A. (2025). The effect of using MatGPT on mathematical proficiency among undergraduate students. International Journal of Information and Education Technology, 15(4), 782–794. https://doi.org/10.18178/ijiet.2025.15.4.2284

Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. ZDM – Mathematics Education, 56(2), 281–292. https://doi.org/10.1007/s11858-023-01530-2

Ergene, O., & Caylan-Ergene, B. (2026). Pre-service mathematics teachers' and engineering students' perceptions of ChatGPT in mathematics: Development, validation and implementation study. Digital Experiences in Mathematics Education, 12, 68–94. https://doi.org/10.1007/s40751-025-00176-x

Etcuban, J. O. (2025). The use of ChatGPT in addressing algebra anxiety and promoting confidence. International Electronic Journal of Mathematics Education, 20(2), em0825. https://doi.org/10.29333/iejme/16008

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Filiz, A., & Gür, H. (2025). Students' perceptions and applications of metacognitive awareness levels in problem solving with ChatGPT. Educational Process International Journal, 14. https://doi.org/10.22521/edupij.2025.14.63

Güler, M., Bütüner, S. Ö., Danişman, Ş., & Gürsoy, K. (2022). A meta-analysis of the impact of mobile learning on mathematics achievement. Education and Information Technologies, 27, 1725–1745. https://doi.org/10.1007/s10639-021-10640-x

Hasratuddin, Dewi, I., Siregar, B. H., & Suwanto. (2026). Mathematical reasoning ability in geometry learning based on a realistic approach assisted by AI for students of the mathematics education study program in State University of Medan. Journal of Cultural Analysis and Social Change, 11(1), 12–22. https://doi.org/10.64753/jcasc.v11i1.3031

Hendriani, M., Arnawa, I. M., & Zainil, M. (2025). Global research trends in mathematical problem-solving

skills: A bibliometric review of 2020–2024 publications. Seminars in Medical Writing and Education, 4, Article 453. https://doi.org/10.56294/mw2025453

Hidayah, N., & Sodikin, S. (2026). Explainable AI, academic confidence, and decision-making quality in physics learning. Digital Learning, Social Science, and Life-course Studies, 2(1), 28–45. https://doi.org/10.70211/disolife.v2i1.777

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15(1), Article 17458. https://doi.org/10.1038/s41598-025-97652-6

Lagon, H. M. (2025). Integrating ChatGPT in math education: A narrative inquiry into mathematical investigation, problem solving, and mathematical modeling. Asian Journal of Education and Social Studies, 51(6), 416–425. https://doi.org/10.9734/ajess/2025/v51i62004

Lau, C. T. G., Tang, H. E., & Chien, T. M. (2025). Exploring foundation students' AI math tool preferences, duration of use, and confidence in mathematics problem-solving. Journal of Mathematics and Science Teacher, 5(4). https://doi.org/10.29333/mathsciteacher/16869

Li, K. (2025). Factors influencing learners' learning interest when using AI chatbots-assisted math learning in higher education. Frontiers in Psychology, 16, Article 1716913. https://doi.org/10.3389/fpsyg.2025.1716913

Lin, W., & Jiang, P. (2025). Factors influencing college students' generative artificial intelligence usage behavior in mathematics learning: A case from China. Behavioral Sciences, 15(3), Article 295. https://doi.org/10.3390/bs15030295

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), Article 410. https://doi.org/10.3390/educsci13040410

Mairing, J. P., Rizaldi, M., Pandiangan, P., Lada, E. Y., & Monita, D. M. (2024). Development of problem, YouTube, and ChatGPT learning model to improve students' proving ability in real analysis. Kreano, Jurnal Matematika Kreatif-Inovatif, 15(1), 169–184. https://doi.org/10.15294/sgz61469

Maknun, C. L., Rosjanuardi, R., & Kusumah, Y. S. (2026). Co-designing Python-based interactive media for trigonometry learning using ChatGPT: A preliminary study. Jurnal Pendidikan Matematika dan IPA, 17(1), 105–117. https://doi.org/10.26418/jpmipa.v17i1.96973

Mardianti, D., Kusimawati, I., & Arifin, B. (2026). Reorienting teacher and lecturer competencies in the era of deep learning and generative artificial intelligence in the transformation of 21st-century learning. JPI: Jurnal Pustaka Indonesia, 6(1), 414–432. https://doi.org/10.62159/jpi.v6i1.2257

Matzakos, N., & Moundridou, M. (2025). Integrating LLMs into mathematics laboratories alongside CAS. International Journal of Engineering Pedagogy, 15(3), 4–18. https://doi.org/10.3991/ijep.v15i3.53151

Matzakos, N., Doukakis, S., & Moundridou, M. (2023). Learning mathematics with large language models: A comparative study with computer algebra systems and other tools. International Journal of Emerging Technologies in Learning, 18(20), 51–71. https://doi.org/10.3991/ijet.v18i20.42979

May, B. M. (2024). Retrieval practice and test-potentiated learning: A comparison of high and low prior topic knowledge students in terms of procedural fluency and conceptual understanding. Pythagoras, 45(1), Article a755. https://doi.org/10.4102/pythagoras.v45i1.755

Mokgwathi, M. S., & Ngoveni, M. A. (2026). Cultivating responsible AI use through formative assessment: Insights from ODeL mathematics students. International Journal of Learning, Teaching and Educational Research, 25(2), 742–762. https://doi.org/10.26803/ijlter.25.2.33

Moral-Sánchez, S. N., Ruiz Rey, F. J., & Cebrián-de-la-Serna, M. (2023). Analysis of artificial intelligence chatbots and satisfaction for learning in mathematics education. IJERI: International Journal of Educational Research and Innovation, 20, 1–14. https://doi.org/10.46661/ijeri.8196

Mustapa, Z., & Masaniku, R. (2025). Photomath and ChatGPT-4o applications in fraction learning: A review from the perspective of mathematics education students. ETDC: Indonesian Journal of Research and Educational Review, 4(3), 805–812. https://doi.org/10.51574/ijrer.v4i3.3691

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

Pala, C. A., Lagrimas, D., & Nobis, M. L., Jr. (2025). Exploration of chatbots in mathematics education for innovative learning process. Asian Journal of Advanced Research and Reports, 19(5), 178–194. https://doi.org/10.9734/ajarr/2025/v19i51010

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. TechTrends, 69, 1105–1120. https://doi.org/10.1007/s11528-025-01100-1

Serhan, D., & Welcome, N. (2024). Integrating ChatGPT in the calculus classroom: Student perceptions. International Journal of Technology in Education and Science, 8(2), 325–335. https://doi.org/10.46328/ijtes.559

Sharma, S., Mittal, P., Kumar, M., & Bhardwaj, V. (2025). The role of large language models in personalized learning: A systematic review of educational impact. Discover Sustainability. https://doi.org/10.1007/s43621-025-01094-z

Sofroniou, A., Patel, M. H., Premnath, B., & Wall, J. (2025). Advancing conceptual understanding: A meta-analysis on the impact of digital technologies in higher education mathematics. Education Sciences, 15(11), Article 1544. https://doi.org/10.3390/educsci15111544

St Omer, S. M., Evers, K., Wang, C.-Y., & Chen, S. (2025). Technology-enhanced mathematics learning: Review of the interactions between technological attributes and aspects of mathematics education from 2013 to 2022. Humanities and Social Sciences Communications, 12, Article 1079. https://doi.org/10.1057/s41599-025-05475-7

Sucipto, L., & Negara, H. R. P. (2023). Assessment of teaching effectiveness, digital literacy, and feedback on student assignments using ChatGPT. Ulul Albab: Majalah Universitas Muhammadiyah Mataram, 27(2), 21–31. https://doi.org/10.31764/jua.v27i2.23330

Supriyadi, E., Luritawaty, I. P., Mutoffar, M. M., & Pahmi, S. (2025). Perceptions of undergraduate mathematics students on generative AI as an algebra lesson planning tool. Journal of Engineering Science and Technology, 20(Special issue), 68–75.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, Article 15. https://doi.org/10.1186/s40561-023-00237-x

Urhan, S., Gençaslan, O., & Dost, Ş. (2025). The role of ChatGPT in supporting students' use of representations: Insights from a case study on calculus concepts. International Journal for Technology in Mathematics Education, 32(4), 181–195. https://doi.org/10.1564/tme_v32.4.03

Urhan, S., Gençaslan, O., & Dost, Ş. (2026). Is ChatGPT a rational assistant for university students during mathematical reasoning? Science & Education. Advance online publication. https://doi.org/10.1007/s11191-026-00725-9

Vankúš, P. (2024). Generative artificial intelligence on mobile devices in the university preparation of future teachers of mathematics. Interactive Mobile Technologies, 18(18), 19–33. https://doi.org/10.3991/ijim.v18i18.51221

Wu, F., Dang, Y., & Li, M. (2025). A systematic review of responses, attitudes, and utilization behaviors on generative AI for teaching and learning in higher education. Behavioral Sciences, 15(4), Article 467. https://doi.org/10.3390/bs15040467

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 55(1), 90–112. https://doi.org/10.1111/bjet.13370

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

Yu, H. M., Lee, H., & Kim, S. H. (2026). Integrating ChatGPT in teacher education: Examining early childhood and elementary pre-service teachers' experiences in AI-assisted math lesson planning. Journal of Research in Childhood Education, 40(1), 67–89. https://doi.org/10.1080/02568543.2025.2579668

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0

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(1), Article 28.

https://doi.org/10.1186/s40561-024-00316-7
Zhuang, Y. (2025). Lessons from using ChatGPT in calculus: Insights from two contrasting cases. Journal of Formative Design in Learning, 9(1), 25–35. https://doi.org/10.1007/s41686-025-00098-2