SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 1.1, (2026), pp. 3490-3502
Open Access. Online & Print

DOI: 10.5281/zenodo.121126255

ARTIFICIAL INTELLIGENCE IMPLEMENTATION IN ENHANCING MATHEMATICAL SKILLS: SYSTEMATIC LITERATURE REVIEW

Thassya Novanka Efendi[1], Mufarrido Husnah[2], Ani Nurhayati[3], Annisa Fauziyyah[4], Noviana Antaria[5], Jumiyati Yamin6, Sarwan[7], Yosi Cahyaningtyas Fitri[8], Pebi Pitri Anasari[9], Riska Indah Lestari[10], Yuce Sandra[11]

[1]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: thassyanovankaefendi@upi.edu
[2]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: mufarrido.husnah@upi.edu
[3]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: aninurhayati@upi.edu
[4]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: annisa.fauziyyah28@upi.edu
[5]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: novianaantaria@upi.edu
[6]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: jumiyatiyamin@upi.edu
[7]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: sarwan@upi.edu
[8]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: yossicahya.student@upi.edu
[9]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: pebipitri@upi.edu
[10]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: riskaindahlestari@upi.edu
[11]Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia. Email: yuce.sandra.student@upi.edu

Received: 15/10/2025
Accepted: 15/01/2026

Corresponding Author: Thassya Novanka Efendi (thassyanovankaefendi@upi.edu)

ABSTRACT

The integration of Artificial Intelligence (AI) into mathematics education has evolved from experimental innovation to a transformative force. Yet, the specific alignment between AI technologies and distinct strands of mathematical proficiency remains fragmented. This systematic literature review aims to synthesize current trends, methodological approaches, and the efficacy of AI tools in enhancing mathematical skills. In accordance with PRISMA guidelines, 10 empirical studies published between 2020 and 2025 were selected from

the Scopus database. The analysis reveals that the field is a nascent research domain, with a surge in publications from 2023 to 2024. Methodologically, qualitative inquiries (50%) and quantitative studies (40%) predominate, focusing on higher education students and pre-service teachers, indicating a significant gap in K-12 settings and mixed-method designs. The findings highlight that Chatbots(25%) and Intelligent Tutoring Systems (16.67%) are the most implemented technologies, with Generative AI tools like ChatGPT and MathGPT serving as primary catalysts for enhancing problem-solving skills and conceptual understanding. However, the review identified a critical lack of discussion of ethics and teacher roles (4.17%), suggesting a techno-centric bias in the current literature. While AI offers promising avenues for personalized scaffolding and democratized tutoring, challenges remain regarding accuracy and student overreliance. This study concludes that future research must pivot towards longitudinal, mixed-method studies and the development of AI literacy programs to ensure ethical and effective integration."

KEYWORDS: Artificial Intelligence, Mathematics Education, Mathematical Skills.

1. INTRODUCTION

Mathematics is a foundational discipline that shapes not only specific innovations but also the practical aspects of modern society, including engineering, finance, technology, and everyday decision-making (Yadav, 2019). The logical structure of mathematics fosters critical thinking, problem-solving abilities, and analytical reasoning, which are necessary for personal and societal advancement (Simonovic et al., 2023). As society becomes more complex, the relevance of mathematics education in empowering individuals to engage with and solve real-world problems continues to grow (Thanheiser, 2023; Lee et al., 2023).

To navigate this complexity, learners must master a broad spectrum of mathematical skills, ranging from number sense, arithmetic, and algebraic thinking to logical reasoning, spatial understanding, measurement, data analysis, and problem-solving (Dreyfus et al., 2018). These skills develop progressively and align with the core components of mathematical proficiency: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition (Kilpatrick et al., 2001). Mastery of these elements is central to the core competencies needed for success in school and beyond, enabling learners to interpret mathematical information effectively and apply appropriate strategies across diverse contexts (Oco et al., 2023).

However, achieving such competence requires a pedagogical approach that goes beyond mere conceptual and procedural knowledge; it demands the ability to select and implement effective strategies. In this context, the integration of AI in education has rapidly progressed from experimental innovation to a transformative force reshaping teaching and learning (Chen et al., 2020). Historically, AI and education have been inextricably intertwined since the 1940s (Mishra et al., 2025). Early pioneers like Newell and Simon (1972) explored fundamental questions about the mechanisms of mental processing and the nature of learning, laying the groundwork for linking cognitive science to educational applications (Hoadley, 2018; Pea, 2016; Doroudi, 2023). Building on these cognitive foundations, modern AI tools have evolved into sophisticated systems capable of addressing educational problems.

AI technologies in mathematics education primarily include Intelligent Tutoring Systems (ITS), Generative AI (GenAI) tools such as ChatGPT and ErnieBot, and Adaptive Learning Systems (ALS) such as Khan Academy. Each technological category presents specific pedagogical benefits. GenAI holds the promise of personalized learning and administrative automation, while ALS utilizes algorithms to assess student progress and tailor instructions to specific learning needs (Alabdulaziz, 2021; Mishra et al., 2025; Orlando & Attard, 2016; Shin, 2022; Wu et al., 2023; Zhao et al., 2025). Essentially, implementing AI in education involves integrating diverse tools into the learning process to provide tailored feedback and address individual learning gaps (Egunjobi & Adeyeye, 2024).

Recent empirical studies highlight the potential of these innovations. AI-driven systems provide scalable solutions for diverse educational needs, enhancing students' conceptual understanding, engagement, and proficiency (Liu et al., 2025). Specifically, AI tools have been shown to advance problem-solving skills and improve overall learning outcomes by offering individualized feedback (Alvarez, 2024; Hwang & Tu, 2021; Marta et al., 2023; Popenici & Kerr, 2017; Vaerenbergh & Pérez-Suay, 2022). Furthermore, GenAI tools are particularly noted for making abstract ideas more accessible, thereby deepening conceptual understanding (Cotič et al., 2024; Zhao et al., 2025). Despite the growing body of research, the literature on the use of AI in mathematics education remains fragmented, with inconsistent findings regarding which specific AI tools most effectively enhance particular mathematical skills. Although numerous studies indicate that AI contributes to students' mathematical competencies, a comprehensive synthesis explaining how different types of AI Influence distinct strands of mathematical proficiency is still lacking. Accordingly, this study seeks to address this gap by systematically synthesizing relevant literature to provide an integrated understanding of how AI is implemented to enhance students' mathematical skills. To accomplish this objective, the study examines the current trends and characteristics of research on AI use in mathematics learning, identifies the specific mathematical skills improved through AI implementation, and analyzes the types of AI technologies employed, explores how AI is implemented within educational settings for mathematics instruction, evaluates the associated benefits and challenges, and outlines potential directions for future research in AI-supported mathematics education.

2. METHOD

The primary objective of this study is to provide a comprehensive analysis of AI implementation to enhance mathematical skills in mathematics education through a systematic literature review. According to Petticrew and Roberts (2006), a systematic literature review is essential for synthesizing large volumes of data to evaluate the effectiveness of educational interventions and to highlight existing research gaps. The review process was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines to ensure transparency, rigor, and reproducibility (Page et al., 2021). The methodology encompasses five distinct phases: (1) protocol development, (2) literature search strategy, (3) eligibility criteria definition, (4) study selection and quality assessment, and (5) data extraction and synthesis.

2.1 Data Sources and Search Strategy

To capture a comprehensive range of relevant studies, a systematic search was conducted across the Scopus database. The Scopus database was selected for its high impact factor and extensive coverage of educational technology literature. The search strategy utilized a combination of keywords related to three main constructs: (1) AI, (2) Mathematics Education, and (3) Mathematical Skills. Boolean operators (AND, OR) were employed to refine the search results. The search string used was adapted for each database's syntax. The general search string was formulated as follows:

TITLE-ABS-KEY (("artificial intelligence" OR "AI" OR "AI TOOL" OR "AI APP" OR "AI Application" OR "AI Implementation") AND ("Mathematics Education") AND ("Problem-Solving" OR "Problem Solving") OR ("Critical Thinking") OR ("Creative Thinking") OR ("Reasoning") OR ("Understanding") OR ("analysis" OR "Analyze") OR ("Representation") OR ("Argumentation") OR ("HOTS" OR "Higher Order Thinking Skills") OR ("Communication") OR ("Conceptual")) AND PUBYEAR > 2020 AND PUBYEAR < 2025.

The literature search was limited to studies published between 2020 and 2025 to ensure the inclusion of the most recent technological advancements, particularly the emergence of GenAI.

2.2 Eligibility Criteria

To ensure the relevance and quality of the included studies, specific inclusion and exclusion criteria are established as detailed in Table 1.

Table 1: Inclusion and Exclusion Criteria.

Criterion	Inclusion Criteria	Exclusion Criteria
Availability	Open access/ full text available	Restricted access/ abstract only/ full text unavailable
Document Type	Empirical research articles (quantitative, qualitative, and mixed methods)	Review articles, book chapters, conference articles, bibliometrics articles, editorials
Language	English	Non-English
Intervention	Implementation of AI tools	Non-AI digital tools
Outcome	Measurement of mathematical skills, proficiency, or cognitive learning outcomes	Focus solely on attitudes or perceptions, without assessing skills

2.3 Study Selection Process

The selection process was conducted following the PRISMA guidelines. A total of 101 records were initially identified from the Scopus database. The complete flow of

the selection process, including the number of records excluded at each stage, is documented in the PRISMA flow diagram as shown in Figure 1.

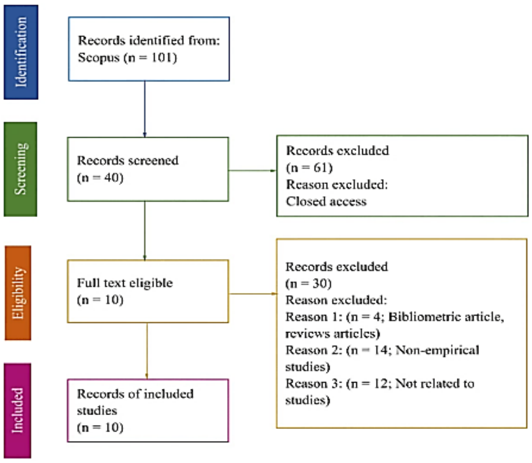

Figure 1: PRISMA Flow Diagram.

In the preliminary phase, 61 records were initially identified from the Scopus database. In the preliminary phase, 61 records were excluded due to closed access, leaving 40 records to be screened. These remaining records underwent a rigorous assessment against the inclusion criteria. During this stage, 30 records were excluded for specific reasons: 14 studies were excluded as they did not present primary data or employ empirical research approaches; 12 studies were removed because they did not utilize AI in the intervention or did not examine the application of AI in enhancing mathematical skills; and 4 studies were identified as bibliometric or review articles. Consequently, 10 full-text articles satisfied all eligibility criteria and were included in the final review.

2.4 Quality Assessment

To ensure the methodological rigor of the synthesized findings, the included studies underwent a quality assessment. The Mixed Methods Appraisal Tool (MMAT)

version 201 was utilized (Hong et al., 2018). The MMAT is designed to appraise the quality of empirical studies, including qualitative, quantitative, and mixed-methods research, making it suitable for the diverse methodologies anticipated in this review. Studies were evaluated based on criteria such as the clarity of research questions, the appropriateness of data collection, and the rigor of analysis.

2.5 Data Extraction and Synthesis

Data from the selected studies were extracted using a structured Data Extraction Form developed specifically for this review. To ensure consistency, the extraction protocol was pilot-tested on a random sample of 5 papers before full implementation. The extracted data items were mapped directly to the research questions and the analysis methods, as detailed in Table 2.

Table 2: Data Extraction Protocol and Analysis Mapping.

Data Category	Specific Items Extracted	Relevant RQ	Analysis Method
Study Characteristics	Author(s), year of publication, participants type	RQ1	Descriptive statistics
Methodology	Research design (quantitative, qualitative, mixed)	RQ1	Descriptive statistics
Targeted Skills	Specific mathematical domains and proficiency strands	RQ2	Thematic analysis
AI Intervention	Type of technology (ITS, GenAI, ALS), name of the tool	RQ3	Content analysis
Implementation Context	Role of AI, pedagogical strategy used	RQ4	Content analysis
Key Findings	Reported benefits, challenges and limitations, statistical outcomes (for quantitative studies), qualitative themes	RQ5	Thematic synthesis
Future Directions	Recommendations for future research, identified gaps	RQ6	Narrative synthesis

Aligning with the data extraction methods, a mixed-methods synthesis approach was adopted. Quantitative data, such as publication trends and methodological distribution (RQ1), were analyzed using descriptive statistics (frequencies and percentages). Qualitative data regarding AI types, mathematical skills, and implementation strategies (RQ2-RQ6) were analyzed using thematic analysis. This involved coding the extracted text, grouping codes into descriptive themes, and interpreting these themes to generate answers to the research questions.

3. RESULTS AND DISCUSSION

3.1 Current Trends and Characteristics in Studies on AI Tools for Enhancing Students' Mathematical Skills

3.1.1 Publication Trends

Analyzing the temporal distribution of research is essential for identifying emerging trends and assessing the maturity of a specific domain. It helps researchers understand the trajectory of technological interaction and anticipate future educational shifts. Although the search scope encompassed 2020-2025, the eligible documents were identified exclusively within 2023-2024. It has been found that the majority of studies (70%; n=7) were published in 2024, with the remaining 30% (n=3) published in 2023. Therefore, Figure 2 presents the annual distribution of selected publications to illustrate this recent surge. In this context, the recent surge in studies indicates that integrating AI to enhance students' mathematical skills is a nascent research trend that has only recently gained significant momentum.

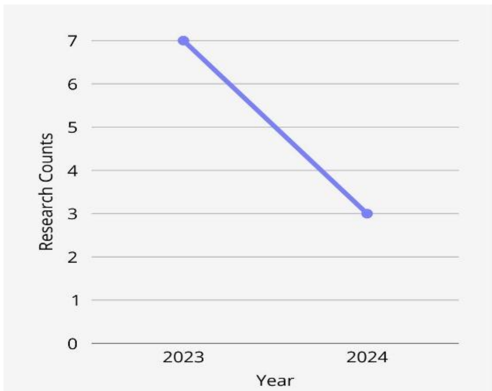

Figure 2: Temporal Distribution of Selected Studies.

3.1.2 Research Methods

The choice of research methodology reflects how scholars currently approach the complexities of integrating AI into mathematics education. Figure 3 illustrates the distribution of research methods utilized in the selected studies. The analysis reveals that qualitative methods are the predominant approach, accounting for 50% of the studies. This dominance suggests that the current research landscape prioritizes exploring how students interact with AI tools, their perceptions, and the pedagogical nuances of explaining mathematical concepts through dialogue. Researchers appear focus on understanding the nature of Human-AI interaction rather than merely measuring output.

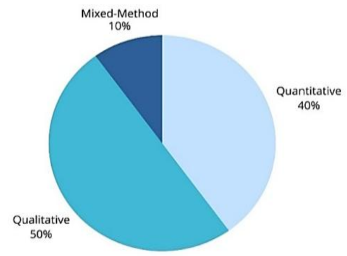

Figure 3: Distributions of Research Methodologies Utilized in the Reviewed Studies.

On the other hand, quantitative methods constitute 40% of the research. This significant portion indicates a parallel interest in empirically validating the efficacy of AI tools through measurable outcomes, such as pretest and posttest scores on mathematical proficiency. However, mixed-method approaches remain the least utilized, accounting for only 10% of the studies. This scarcity highlights a potential methodological gap. While qualitative studies provide depth and quantitative studies provide breadth, the limited use of mixed-method designs suggests that few studies are currently triangulating statistical data with contextual insights to provide a holistic view. Future research would benefit from using mixed-methods designs to more robustly validate statistical gains in relation to student experiences.

3.1.3 Participant Type

Analyzing the target population is crucial for understanding the context in which AI interventions are being tested. As illustrated in Figure 4, the distribution of participant types reveals a concentration of research within the tertiary education sector. Students in Higher Education and Pre-Service Teachers emerged as the most frequently investigated groups, each represented in 3 studies (30% each). This focus on university-level participants (Higher Education and Pre-Service Teachers) mirrors trends in educational technology research, where convenience sampling plays a vital role. University researchers have easier access to undergraduate students and teacher candidates than to K-12 classrooms. Furthermore, focusing on Pre-Service and Teachers aligns with the findings of Listiaji and Molnár (2025), who noted that integrating computational tools at this stage is strategic for preparing future educators before they enter the workforce.

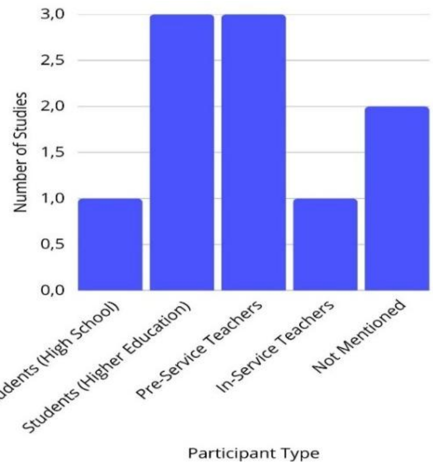

Figure 4: Classification of Study Participants.

In contrast, research involving high school students and in-service teachers is limited, with only 1 study identified for each group. Additionally, a notable portion of studies (n = 2) did not explicitly specify the participant type. The underrepresentation of K-12 students suggests a lack of empirical evidence on how AI tools function in actual school settings with younger learners. While higher education students may possess the digital literacy to navigate AI tools like ChatGPT effectively, younger students may require different pedagogical scaffolds. Therefore, future research should prioritize conducting empirical studies directly in K-12 classrooms and in-service teachers to ensure that AI integration strategies are ecologically valid for school environments.

3.2 Specific Mathematical Skills Enhanced through AI Implementation

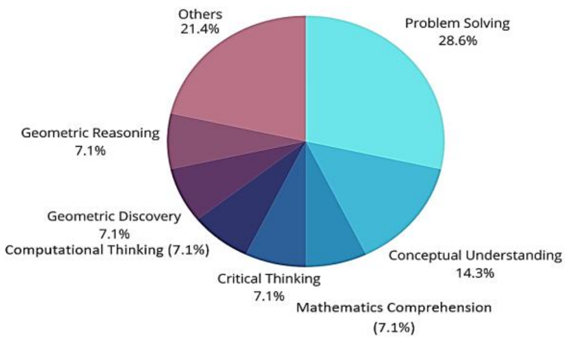

Figure 5: Distribution of Mathematical Skills Targeted by AI Interventions.

As illustrated in Figure 5, the scope ofmathematical skills investigated in the selected studies remains relatively concentrated. The analysis identified specific proficiencies, including critical thinking, problem-solving, computational thinking, comprehension, and geometric reasoning. Among the mathematical skills mentioned, problem-solving skills emerged as the most frequently examined competency, followed by conceptual understanding. Several studies adopted a multi-dimensional approach, targeting more than one skill simultaneously. For instance, Alvarez (2024) demonstrated that AI-powered tutors, such as MathGPT and Flexi 2.0, significantly enhanced students' problem-solving abilities and comprehension in Calculus. Similarly, Wardat et al. (2023) examined stakeholders' perspectives on the utility of ChatGPT in improving these same two domains. In geometry, Botana et al. (2024) investigated the use of ChatGPT for automated geometric reasoning and discovery.

Beyond core mathematical proficiencies, the literature also addresses broader competencies and affective dimensions associated with AI integration. Bagno et al. (2024) used ChatGPT in linear algebra instruction not only to develop subject-matter expertise but also to foster critical thinking and curiosity. Furthermore, Moral-Sánchez et al. (2023) examined the role of AI chatbots in enhancing digital competence, while Li (2024) shifted the focus to pedagogical enablers, exploring mathematics teachers' self-efficacy in technology integration. Collectively, these findings indicate that while AI implementation is growing, the distribution of mathematical skills measured in current studies remains relatively uniform, with a strong emphasis on problem-solving and conceptual understanding.

3.3 Types of AI Technologies Utilized to Enhance Mathematics Skills

Analyzing the alignment between specific AI technologies and targeted mathematical skills is essential for understanding the pedagogical utility of each tool. Table 3 presents a synthesis of the included studies, categorizing how distinct AI types, ranging from large language models to specialized discovery engines, are deployed to enhance specific cognitive and affective competencies. The analysis reveals a clear trend, while Generative AI (GenAI) serves as a versatile tool for broad conceptual understanding, specialized tools are often reserved for domain-specific reasoning, such as geometry.

Table 3: Mapping of AI Technologies and Targeted Mathematical Skills.

Types of AI	Mathematical Skills	Author
ChatGPT	Conceptual Understanding, Geometric Reasoning and Discovery, Problem-Solving, Critical Thinking	(Alvarez, 2024; Bagno et al., 2024; Botana et al., 2024; Getenet, 2024; Torres-Peña et al., 2024; Wardat et al., 2023; Yunianto et al., 2024)
Chatbot	Problem-Solving, Digital Competence	(Jančařík et al., 2023; Moral-Sánchez et al., 2023)
MathGPT	Mathematical Comprehension, Problem-Solving, Conceptual Understanding	(Alvarez, 2024; Torres-Peña et al., 2024)
GeoGebra Discovery	Geometric Reasoning and Discovery	(Botana et al., 2024)
Gemini, Wolfram Alpha, Flexi 2.0	Conceptual Understanding	(Alvarez, 2024; Torres-Peña et al., 2024)
Others	Self-Efficacy for Technology Integration	(Li, 2024)

3.3.1 Generative AI as a Catalyst for Conceptual and Problem-Solving Skills

The findings indicate that GenAI tools, specifically ChatGPT, MathGPT, Gemini, and Flexi 2.0, are predominantly utilized to foster conceptual understanding and problem-solving skills (Alvarez, 2024; Bagno et al., 2024; Torres-Peña et al., 2024; Yunianto et al., 2024). As noted in the analysis, ChatGPT and MathGPT are frequently associated with enhancing mathematical comprehension and critical thinking (Alvarez, 2024; Torres-Peña et al., 2024; Wardat et al., 2023). The reviewed research indicates that educators are leveraging GenAI's natural language processing capabilities to demystify complex mathematical concepts.

Unlike traditional calculators that provide only the final answer, these tools offer step-by-step explanations, functioning as personalized tutors that help students bridge the gap between procedural steps and deep conceptual grasp. For instance, Alvarez (2024) highlighted the role of MathGPT in improving calculus comprehension, illustrating how AI can transform abstract symbolic manipulation into understandable logic. This aligns with the adaptive reasoning strand of mathematical proficiency, in which students learn to justify and explain their thinking.

3.3.2 Specialized Tools for Geometric Reasoning and Discovery

In the Geometry domain, the landscape of AI implementation becomes more specialized. The data shows that GeoGebra Discovery and specific applications of ChatGPT are utilized to enhance Geometric Reasoning and Discovery (Botana et al., 2024; Getenet, 2024). This finding highlights a distinct pedagogical need in geometry education: the necessity for visualization and logical deduction. Text-based GenAI, such as ChatGPT, supports reasoning by generating proofs or describing properties. Tools like GeoGebra Discovery provide the dynamic visual environment needed to discover formulas and concepts in Geometry (Botana et al., 2024). The combination of these tools suggests a hybrid approach in which AI supports both visual-spatial exploration and the logical-linguistic articulation required to master geometry. This indicates that, in highly visual mathematical domains, general-purpose AI is often supplemented by or integrated into domain-specific software to achieve optimal results.

3.3.3 Expanding Beyond Cognition: Digital Competence and Self-Efficacy

Beyond cognitive mathematical skills, the analysis reveals that AI tools are also instrumental in developing transversal competencies. As shown in Table 3, Chatbots and other AI tools are linked to enhanced digital competence and self-efficacy in technology integration (Jančařík et al., 2023; Li, 2024; Moral-Sánchez et al., 2023). As AI is not merely a subject-learning tool but an enabler and boundary for 21st-century skills, these findings are a necessary insight. The focus on self-efficacy implies that successful interaction with AI tools can boost students' confidence not only in mathematics but also in their ability to navigate the technological environment (Li, 2024). This resonates with the productive disposition strand of mathematical proficiency (Kilpatrick et al., 2001), in which students perceive mathematics as sensible and useful, and they believe in their ability to solve problems using modern tools. Therefore, the implementation of AI serves a dual purpose: reinforcing mathematical content knowledge while simultaneously building the digital resilience required in the modern era.

3.4 The Use of AI to Support Mathematics Instruction in Education

As depicted in Figure 6, the thematic analysis categorized AI implementation into seven distinct pedagogical domains. Chatbots emerged as the predominant application, constituting 25% of the analyzed studies. This prevalence indicates a significant shift towards interactive, dialogue-based learning environments where AI functions as a conversational agent. Following this, Intelligent Tutoring Systems and Problem-solving and Reasoning Support each accounted for 16.67% of the distribution. These findings underscore the role of AI in delivering adaptive instruction and cognitive scaffolding to support complex mathematical tasks.

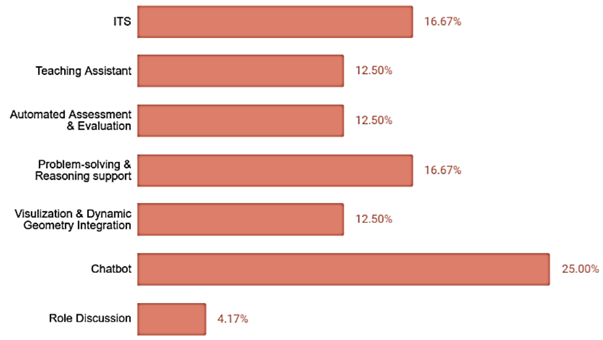

Figure 6: Thematic Categorization of AI Implementation Roles in Mathematics Education.

Furthermore, three categories, such as Teaching Assistant, Automated Assessment and Evaluation, and Visualisation and Dynamic Geometry Integration, represented 12.50% per each of the studies. This reflects the diverse utility of AI, ranging from instructional support and automated grading to the visual representation of abstract mathematical concepts. Conversely, the theme of Role Discussion appeared least frequently (4.17%), suggesting that theoretical discourse on the ethical and policy-level implications of AI integration remains limited relative to empirical applications.

3.4.1 The Shift from Static Tools to Interactive Partners

The findings reveal a clear trajectory in the evolution of AI in mathematics education. The dominance of Chatbots (25%) and ITS (16.67%) suggests that the pedagogical paradigm is shifting from viewing technology as a passive tool to embracing it as an interactive learning partner. This trend aligns with the rapid advancement of GenAI, which has democratized access to conversational tutoring. Unlike traditional software that offers static responses, these AI agents provide personalized, step-by-step feedback, thereby mimicking the role of a human tutor. This is particularly crucial in mathematics education, where immediate feedback on problem-solving processes is essential for misconceptions (Alvarez, 2024; Torres-Peña et al., 2024).

3.4.2 Bridging the Gap in Visualization and Assessment

The moderate distribution of visualization (12.5%) and automated assessment (12.5%) highlights AI's expanding role in addressing specific challenges in mathematics. The use of AI for dynamic geometry integration suggests that educators are leveraging technology to make abstract spatial concepts more concrete, a core component of mathematical proficiency (Yunianto et al., 2024). Meanwhile, the implementation of AI as a Teaching Assistant and for assessment demonstrates a move towards operational efficiency, freeing educators from repetitive grading tasks so they can focus on instructional strategy (Moral-Sánchez et al., 2023).

3.4.3 The Critical Gap in Ethical and Role Discourse

A significant finding is the underrepresentation of Role Discussion (4.17%). The scarcity of studies focusing on the ethical implications, policy frameworks, and the changing role of teachers suggests a potential gap in the current research landscape. While the field is heavily focused on technical implementation and efficacy, there is a lag in critically examining the broader consequences of AI adoption. This indicates a critical need for future research to balance empirical trials with theoretical inquiries into how AI redefines the teacher-student dynamic and the ethical boundaries of automated tutoring in mathematics classrooms (Alvarez, 2024; Bagno et al., 2024).

3.5 Benefits and Challenges Associated with AI Implementation in Enhancing Mathematics Skills

Findings from this study reveal a clear dichotomy in the integration of AI tools, such as MathGPT, Flexi 2.0, and various large language models Chatbots, for enhancing mathematical skills. On one hand, the findings overwhelmingly support the benefits of AI primarily through its capacity for personalization and adaptive learning. Students using AI tutors demonstrate significant improvements in problem-solving and advanced cognitive abilities compared to conventional instruction, benefiting from immediate feedback and a learning environment tailored to their individual needs (Alvarez, 2024; Jančařík et al., 2023; Torres-Peña et al., 2024). Furthermore, AI can play a powerful pedagogical role by supporting the development of Computational thinking skills, including programming and debugging, when paired with tools such as GeoGebra (Botana et al., 2024). The dynamic and adaptive nature of AI also substantially boosts students' engagement and motivation, offering uninterrupted support that facilitates mitigating educational inequalities by increasing access to private supplementary tutoring (Jančařík et al., 2023; Torres-Peña et al., 2024). On the other hand, several critical challenges temper this optimism, necessitating cautious implementation. A major limitation is the inherent deficiencies of current AI models, which often fail at logical inference and make significant errors in advanced topics such as linear algebra and geometry, as they rely more on statistical patterns than on genuine mathematical understanding (Botana et al., 2024; Getenet, 2024; Li, 2024). This lack of deep conceptual understanding means that AI cannot effectively correct students' misconceptions, raising concerns about the accuracy and reliability of its outputs (Getenet, 2024). The most pervasive challenge in implementing AI in mathematics education centres on pedagogical integrity and student overreliance. Studies highlight the potential for AI to impact the development of critical thinking negatively, as students may passively receive AI-generated responses rather than actively engage in the learning process (Alvarez, 2024; Botana et al., 2024; Torres-Peña et al., 2024) Pragmatic obstacles also exist, including organizational issues, the need for rigorous implementation criteria, and the necessity fir training students and staff to ethically and effectively use these tools (Alvarez, 2024; Li, 2024). Therefore, AI is a powerful augmentative tool, but not a replacement for human educators (Alvarez, 2024; Jančařík et al., 2023). Successful integration demands a shift in teaching practice, moving the educators' role from content delivery to the critical evaluation and justification of AI-generated results (Alvarez, 2024; Bagno et al., 2024). The successful implementation of AI in mathematics education hinges on designing structured learning activities that explicitly address the risks of overreliance and academic dishonesty, ensuring that ethical considerations and teacher oversight are paramount for maximizing AI's advantages while mitigating its significant challenges.

3.6 Future Research Directions for AI Implementation in Mathematics Education

The reviewed literature identifies several emerging directions for future research, encompassing pedagogical integration, methodological improvement, technological collaboration, teacher capacity building, cross-cultural investigation, human–AI collaboration, critical evaluation of challenges, and the broader generalization of contextual findings. From a pedagogical perspective, Alvarez (2024) emphasized the importance of prioritizing the integration of AI tutors, such as MathGPT and Flexi, into instructional methodologies rather than focusing solely on individual AI tools. This shift highlights the role of AI in complementing conventional teaching approaches, particularly in domains that require complex problem-solving skills. Similarly, Gatenet (2024) argued that teachers should develop sufficient AI literacy to effectively incorporate conversational AI, such as ChatGPT, as a multipurpose learning support tool that extends beyond algebraic problem-solving. Methodologically, Torres-Peña et al. (2024) proposed the need for longitudinal research designs to evaluate the sustained impact of AI implementation on students' learning outcomes and to assess the scalability of AI integration across diverse educational contexts. Wardat et al. (2023) further underscored the value of longitudinal and multilingual approaches to capture broader patterns of AI adoption beyond early users, while also recommending expanded participant samples and the integration of qualitative and quantitative methods to enhance the generalizability of findings. In terms of technological development, Botana et al. (2024) highlighted the potential of collaboration between conversational AI systems and computational mathematics tools. Specifically, integrating chatbots such as ChatGPT with GeoGebra Discovery in Euclidean geometry could enable logical verification of AI-generated proofs, the issuance of proof certificates, and cross-validation across multiple computational environments, thereby strengthening the rigour and transparency of AI-assisted mathematical reasoning. Teacher capacity building also emerged as a critical research direction. Li (2024) and Yunianto et al. (2024) emphasized the need for continuous professional development and structured AI literacy training for educators. Recommendations include periodic updates to the curriculum and assessments to remain aligned with rapid AI advancements, and the integration of AI literacy and data-driven pedagogy within technological pedagogical content Knowledge. framework; the development of prompt engineering literacy; the enhancement of GeoGebra competencies; and the systematic validation of AI-generated outputs as part of teachers' professional growth. Beyond individual classrooms, Li (2024) proposed cross-cultural and multi-level educational investigations to evaluate the universality and contextual variability of AI-supported learning models across educational systems and cultural settings. Such studies would help identify how sociocultural factors influence AI adoption and effectiveness in mathematics education. Future research is also expected to explore human–AI collaborative tutoring models. Wardat et al. (2023) suggested examining how human tutors and conversational AI can function synergistically to achieve learning objectives, whereas Jančařík et al. (2023) emphasized the potential of AI to provide accessible and cost-free tutoring while supporting the design of adaptive problem sets that enhance course effectiveness. At the same time, critical examination of challenges and potential negative impacts remains essential. Bagno et al. (2024) warned that current implementations of ChatGPT in mathematics education remain limited and may lead to adverse outcomes when applied without sufficient pedagogical oversight, highlighting risks such as misinformation, excessive reliance on AI, and ethical concerns in learning environments. Finally, Moral-Sánchez et al. (2023) recommended extending contextual findings to broader populations and educational settings and refining research instruments through expert validation to ensure accuracy, relevance, and adaptability in rapidly evolving technological landscapes. Collectively, these directions underscore the need for integrative, methodologically rigorous, and context-sensitive research to advance the effective and responsible implementation of AI in mathematics education.

4. LIMITATIONS, IMPLICATIONS, AND FUTURE DIRECTIONS

The results of this review should be interpreted within the context of its limitations. The Scopus database's restrictions and English-language publication requirements may have excluded relevant studies from other indices or regions. Additionally, the rigorous inclusion criteria resulted in a small final sample size (n = 10), which means that while ensuring quality, it limits the generalizability of the findings to broader educational contexts.The integration of AI in mathematics education holds transformative potential but demands a cautious approach. For educators, the implication is clear: AI should function as an augmentative tool rather than a replacement. Future research must pivot from exploring what AI can do to investigating how it should be regulated and integrated into K-12 classrooms. Priority should be given to longitudinal mixed-methods studies that evaluate the long-term impact of AI on critical thinking, and to the development of AI literacy programs that empower teachers to navigate the ethical complexities of human-AI collaboration.

5. CONCLUSION

This systematic literature review synthesized empirical evidence from 2020 to 2025 on the implementation of AI to enhance mathematical skills. The analysis of 10 selected studies, predominantly published during 2023 to 2024, confirms that the integration of AI in mathematics education is a nascent yet rapidly accelerating research domain. The review highlights a significant paradigm shift from static technological tools to interactive learning partners, with Chatbots and Intelligent Tutoring Systems accounting for most implementations. The findings reveal that Generative AI tools, particularly ChatGPT and MathGPT are the most utilized technologies, serving as catalysts for enhancing problem-solving skills and conceptual understanding. These tools are lauded for their ability to provide personalized, step-by-step scaffolding, democratizing access to tutoring. However, the pedagogical landscape is currently fragmented. While specialized tools like GeoGebra Discovery are effective for geometric reasoning, general-purpose large language models often lack the precision needed for advanced logical inference, raising concerns about accuracy and students' overreliance. A critical analysis of the researchmethodology indicated a strong preference for qualitative inquiries (50%) focused on higher education students and pre-service teachers, with a notable underrepresentation of K-12 settings and mixed-method designs. Furthermore, the review identified a substantial gap in the literature on roles and ethics. Only 4.17% of studies critically examined the ethical implications or policy frameworks necessary for AI integration. This suggests that current research is heavily techno-centric, prioritizing efficacy over the pedagogical and ethical consequences of automation.

ACKNOWLEDGEMENT

The lead author acknowledges the financial support provided by the Indonesia Endowment Fund for Education (*Lembaga Pengelola Dana Pendidikan*) during the master's research upon which this article is based.

REFERENCES

Alabdulaziz, M. S. (2021). COVID-19 and the use of digital technology in mathematics education. *Education and Information Technologies*, *26*(6), 7609–7633. https://doi.org/10.1007/s10639-021-10602-3

Alvarez, J. I. (2024). Evaluating the impact of AI-powered tutors, MathGPT and FLEXI 2.0, in enhancing calculus learning. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8*(2), 495–508. https://doi.org/10.22437/jiituj.v8i2.34809

Bagno, E., Dana-Picard, T., & Reches, S. (2024). ChatGPT in teaching linear algebra: Strides forward, steps to go. *Open Education Studies*, *6*(1), 1–18. https://doi.org/10.1515/edu-2024-0031

Botana, F., Recio, T., & Vélez, M. P. (2024). On using GeoGebra and ChatGPT for geometric discovery. *Computers*, *13*(8), 1–30. https://doi.org/10.3390/computers13080187

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Cotič, M., Doz, D., Jenko, M., & Žakelj, A. (2024). Mathematics education: What was it, what is it, and what will it be? *International Electronic Journal of Mathematics Education*, *19*(3), em0783. https://doi.org/10.29333/iejme/14663

Doroudi, S. (2023). The intertwined histories of artificial intelligence and education. *International Journal of Artificial Intelligence in Education*, *33*(4), 885–928. https://doi.org/10.1007/s40593-022-00313-2

Dreyfus, T., Artigue, M., Potari, D., Prediger, S., & Ruthven, K. (2018). *Developing research in mathematics education: Twenty years of communication, cooperation, and collaboration in Europe*. London: Routledge

Egunjobi, D., & Adeyeye, O. J. (2024). Revolutionizing learning: The impact of augmented reality (AR) and artificial intelligence (AI) on education. *International Journal of Research Publication and Reviews*, *5*(10), 1157–1170. https://doi.org/10.55248/gengpi.5.1024.2734

Getenet, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: implications for mathematics teaching in primary schools. *International Electronic Journal of Mathematics Education*, *19*(1), 1–12. https://doi.org/10.29333/iejme/14141

Hoadley, C. (2018). *International handbook of the learning sciences* (F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann, Eds.). London: Routledge. https://doi.org/10.4324/9781315617572

Hong, Q. N., Gonzalez-Reyes, A., & Pluye, P. (2018). Improving the usefulness of a tool for appraising the quality of qualitative, quantitative, and mixed methods studies, the mixed methods appraisal tool (MMAT). *Journal of evaluation in clinical practice*, *24*(3), 459–467. https://doi.org/10.1111/jep.12884

Hwang, G.-J., & Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A Bibliometric mapping analysis and systematic review. *Mathematics*, *9*(6), 1–19. https://doi.org/10.3390/math9060584

Jančařík, A., Michal, J., & Novotná, J. (2023). Using a chatbot for math tutoring. *Journal of Education Culture and Society*, *2*(2023), 285–296. https://orcid.org/0000-0003-3331-2396

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Lee, S. W. Y., Tu, H. Y., Chen, G. L., & Lin, H. M. (2023). Exploring the multifaceted roles of mathematics learning in predicting students' computational thinking competency. *International Journal of STEM Education*, *10*(1). 1–16. https://doi.org/10.1186/s40594-023-00455-2

Li, M. (2024). Assessing Chinese primary mathematics teachers' self-efficacy for technology integration: Development and validation of a multifaceted scale. *Asian Journal for Mathematics Education*, *3*(2), 231–253. https://doi.org/10.1177/27527263241254496

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,* 1–11. https://doi.org/10.1016/j.caeai.2025.100438

Marta, P., Khairinal, K., & Murbojono, R. (2023). Innovative approaches: Unveiling integrated social sciences learning at SMP Negeri 7 Muaro Jambi. *Tekno - Pedagogi : Jurnal Teknologi Pendidikan*, *13*(1), 41–49. https://doi.org/10.22437/teknopedagogi.v13i1.32541

Mishra, P., Henriksen, D., Woo, L. J., & Oster, N. (2025). Control vs. agency: Exploring the history of AI in education. *TechTrends*, *69*(2), 247–253. https://doi.org/10.1007/s11528-025-01064-2

Newell, A., & Simon, H. A. (1972). *Human problem solving* (9th ed., Vol. 104). NJ: Prentice-Hall.

Listiaji, P., & Molnár, G. (2025). Integrating computational thinking into teacher education: A systematic literature review. *International Journal of Educational Research, 133*(12), 102682. https://doi.org/10.1016/j.ijer.2025.102682

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

Oco, R. M., Jade, S., & Sabasaje, N. (2023). Mathematical skills and performance. In *International Journal of Multidisciplinary Research and Publications, 6(2),* 328–332.

Orlando, J., & Attard, C. (2016). Digital natives come of age: The reality of today's early career teachers using mobile devices to teach mathematics. *Mathematics Education Research Journal*, *28*(1), 107–121. https://doi.org/10.1007/s13394-015-0159-6

Page et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research ed.)*, *1*(1), 1–13. https://doi.org/10.1136/bmj.n71

Pea, R. D. (2016). The prehistory of the learning sciences. In *Reflections on the Learning Sciences* (pp. 32–58). United Kingdom: Cambridge University Press. https://doi.org/10.1017/CBO9781107707221.003

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Oxford: Blackwell Publishing.

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(1), 1–13. https://doi.org/10.1186/s41039-017-0062-8

Shin, D. (2022). Teaching mathematics integrating intelligent tutoring systems: Investigating prospective teachers' concerns and TPACK. *International Journal of Science and Mathematics Education*, *20*(8), 1659–1676. https://doi.org/10.1007/s10763-021-10221-x

Simonovic, B., Vione, K., Stupple, E., & Doherty, A. (2023). *Thinking Skills and Creativity*, *49*(1), 101362. https://doi.org/10.1016/j.tsc.2023.101362

Thanheiser, E. (2023). What is the mathematics in mathematics education? *Journal of Mathematical Behavior*, *70*(1), 101033. https://doi.org/10.1016/j.jmathb.2023.101033

Torres-Peña, R. C., Peña-González, D., Chacuto-López, E., Ariza, E. A., & Vergara, D. (2024). Updating calculus teaching with AI: A classroom experience. *Education Sciences*, *14*(9), 1–21. https://doi.org/10.3390/educsci14091019

Vaerenbergh, V. S., & Pérez-Suay, A. (2022). *A classification of artificial intelligence systems for mathematics education.* In: Richard, P.R., Vélez, M.P., Vaerenbergh, V. S. (eds). mathematics education in the age of artificial intelligence. Cham: Springer. https://doi.org/10.1007/978-3-030-86909-0_5

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2286. https://doi.org/10.29333/ejmste/13272

Wu, T. T., Lee, H. Y., Wang, W. S., Lin, C. J., & Huang, Y. M. (2023). Leveraging computer vision for adaptive learning in STEM education: Effect of engagement and self-efficacy. *International Journal of Educational Technology in Higher Education*, *20*(1), 1–26. https://doi.org/10.1186/s41239-023-00422-5

Yadav, S. (2019). Role of mathematics in the development of society. *International Journal of Research and Analytical Reviews*, *6*(4), 1–24.

Yunianto, W., Lavicza, Z., Kastner-Hauler, O., & Houghton, T. (2024). Investigating the use of ChatGPT to solve a GeoGebra-based mathematics+computational thinking task in a geometry topic. *Journal on Mathematics Education*, *15*(3), 1027–1052. https://doi.org/10.22342/jme.v15i3.1027-1052