ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 11 Number 3, August 2026, 711-726

THE FUTURE OF MATHEMATICAL LITERACY IN THE ERA OF ARTIFICIAL INTELLIGENCE: A SYSTEMATIC LITERATURE REVIEW

Andi Mangaraja[1*], Dewi Aminah Hasibuan[2], Ramadhan Herianto[3], Henra Putra Nainggolan[4], Sri Ulfah Afriani Dalimunthe[5], Arjuna Yahdil Fauza Rambe[6], Ahmad Nizar Rangkuti[7]

[1,2,3,4,5,6,7]Program Pascasarjana Tadris Matematika, Universitas Islam Negeri Syekh Ali Hasan Ahmad Addary, Sumatera Utara, Padangsidimpuan

*Correspondence: andimangaraja288@gmail.com

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) has transformed educational practices, particularly in mathematics education, by enabling adaptive learning, personalized instruction, and instant feedback. While AI-powered technologies such as ChatGPT, Gemini, and Intelligent Tutoring Systems have been reported to support students' conceptual understanding and learning experiences, concerns have emerged regarding excessive reliance on AI, which may reduce mathematical reasoning, critical thinking, and learning autonomy. Although prior reviews have examined the general effectiveness of AI in education, few have specifically synthesized how AI reshapes mathematical literacy, and none has systematically addressed this issue within the context of Generative AI. This study addresses that gap by systematically examining the role of AI in the development of mathematical literacy, identifying its benefits and challenges, and exploring future research directions in mathematics education. A Systematic Literature Review (SLR) was conducted following the PRISMA 2020 guidelines. Literature was retrieved from Google Scholar, ERIC, Scopus, and Garuda databases using predefined keyword combinations related to AI and mathematical literacy, and screened against explicit inclusion and exclusion criteria. From 287 identified publications, 32 studies published between 2015 and 2024 met the inclusion criteria and were analyzed using thematic analysis. The findings indicate that AI is reported to contribute to mathematical literacy by supporting personalized learning, adaptive feedback, conceptual understanding, and problem-solving skills, although the evidence is not uniformly consistent across contexts, AI tools, and methodological designs. At the same time, excessive AI use may encourage cognitive offloading, reduce independent reasoning, and increase students' dependence on automated solutions. Beyond synthesizing existing evidence, this review proposes that mathematical literacy in the AI era should extend beyond traditional competencies to include the ability to critically evaluate AI-generated outputs, identify algorithmic limitations, and use AI responsibly a construct this review tentatively terms Mathematical AI Literacy. The review concludes by outlining the theoretical contribution of this construct and the empirical gaps that remain, including the absence of standardized Mathematical AI Literacy instruments and the limited number of context-specific studies, particularly in Indonesia.

Keywords: Artificial Intelligence, Mathematical Literacy, Mathematics Education, AI in Education, Systematic Literature Review.

How to Cite: Mangaraja, A., Hasibuan, D. A., Herianto, R., Nainggolan, H. P., Dalimunthe, S. U. A., Rambe, A. Y. F., & Rangkuti, A. N. (2026). The Future of Mathematical Literacy in The Era of Artificial Intelligence: A Systematic Literature Review. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 11*(3), 711-726. http://doi.org/10.31943/mathline.v11i3.1164

PRELIMINARY

The transition from the Fourth Industrial Revolution (Industry 4.0) to the Society 5.0 era has fundamentally transformed various sectors of human life, including education. The integration of digital technology into teaching and learning is no longer regarded as a complementary component but has become an indispensable element of modern educational systems. Among the most influential technological developments in recent years is Artificial Intelligence (AI), which enables computers to emulate a wide range of human cognitive capabilities, including pattern recognition, natural language understanding, logical reasoning, and automated information generation. The emergence of AI has shifted educational paradigms from traditional teacher-centered instruction toward more adaptive, personalized, and data-driven learning environments (Anwari & Nuriadin, 2025).

The advancement of AI has accelerated significantly with the emergence of Generative Artificial Intelligence technologies, such as ChatGPT, Gemini, Claude, Microsoft Copilot, and other Large Language Models (LLMs), which are capable of generating text, solving mathematical problems, explaining scientific concepts, and providing instant learning recommendations. These technologies have been increasingly adopted in education as learning media, virtual assistants, Intelligent Tutoring Systems (ITS), and adaptive learning platforms designed to accommodate individual learning needs, with some scholars arguing that AI is beginning to function not merely as a tool but as an emerging educational platform in its own right (Gadanidis,2017). Unlike conventional educational technologies, AI is reported to provide real-time feedback, identify students' misconceptions, and adjust the difficulty level of learning materials according to learners' abilities, which is argued to enhance the effectiveness and efficiency of the learning process (Febrianti et al., 2025; Kasneci et al., 2023).

Understanding the implications of this shift requires a theoretical foundation broader than technology adoption alone. Mathematics education research has long emphasized mathematical reasoning the capacity to justify, generalize, and connect mathematical ideas as a competency distinct from procedural fluency (English, 2023). At the same time, the increasing mediation of learning by intelligent systems brings mathematical reasoning into direct contact with digital literacy and AI literacy, understood as the ability to use, interpret, and critically evaluate outputs produced by algorithmic systems, as well as with computational thinking as a problem-decomposition skill increasingly identified as part of twenty-first-century competencies (Kasneci et al., 2023). Cognitive load theory further offers an explanatory mechanism for why automation can be a double-edged tool: when AI absorbs part of the germane cognitive load associated with working through a problem, it can either free capacity for higher-order reasoning or, if the schema-building step is bypassed entirely, prevent that reasoning from developing at all (Sweller, 1988). Framing AI's role in mathematics education solely through the PISA definition of mathematical literacy the ability to formulate, employ, and interpret mathematics in context therefore risks overlooking how these adjacent competencies interact with, and potentially redefine, that formulation in the AI era. This concern is consistent with longstanding calls in the field to ground mathematics education research in models that explicitly connect abstract mathematical reasoning to authentic, real-world problem contexts (Lesh, Hamilton, & Kaput, 2020), a connection that becomes only more pressing once AI systems mediate how students encounter those contexts.

Within the context of mathematics education, AI presents substantial opportunities to improve learning outcomes. Mathematics is a discipline that requires logical reasoning, systematic thinking, problem-solving abilities, mathematical communication, and the capacity to connect abstract concepts with real-world situations. Previous studies suggest that AI facilitates students' understanding of complex mathematical concepts through dynamic visualization, interactive simulations, step-by-step problem-solving guidance, and personalized practice activities tailored to students' levels of mastery (Ma et al., 2014; Wardat et al., 2023). Technologies such as Wolfram Alpha, AI-powered GeoGebra, ChatGPT, and Intelligent Tutoring Systems enable learners to experience more individualized learning compared with conventional instructional approaches.

Despite these opportunities, the existing body of research on AI and mathematics education remains fragmented in ways that motivate the present review. Most prior studies have investigated AI's effectiveness in general learning outcomes achievement, motivation, or technology acceptance while comparatively few have examined its specific influence on mathematical literacy as a construct. Existing systematic reviews of AI in education, such as Zawacki-Richter et al. (2019), have similarly emphasized educational technology adoption broadly rather than discipline-specific competencies such as mathematical literacy. Moreover, no recent synthesis has comprehensively evaluated AI's influence on mathematical literacy specifically within the era of Generative AI, where the nature of AI-generated output (fluent, plausible, but not always correct) differs qualitatively from earlier rule-based ITS. This gap in the literature is the primary justification for the present review.

The rapid advancement of AI consequently raises fundamental questions regarding the evolving meaning of mathematical literacy in the age of artificial intelligence. As AI systems become capable of solving increasingly sophisticated mathematical problems automatically, the competencies required of learners extend beyond computational skills. Students must also be able to evaluate the validity of AI-generated responses, identify algorithmic errors, verify solution procedures, select the most appropriate solutions, and utilize AI critically, ethically, and responsibly (Rayhan & Juandi, 2023). Accordingly, mathematical literacy in the AI era encompasses not only mathematical reasoning but also the ability to interact intelligently with AI-driven technologies.

Consistent with these developments, research on AI applications in mathematics education has expanded considerably in recent years, but the findings across this body of work are not consistent. On one hand, a number of studies report that AI contributes positively to students' academic achievement, learning motivation, personalized instruction, feedback efficiency, and conceptual understanding of mathematics; for instance, Ma et al. (2014) found in a meta-analysis that Intelligent Tutoring Systems improve academic performance through adaptive learning, while Wardat et al. (2023), using an experimental design, reported that ChatGPT broadens learners' conceptual understanding by providing explanations from multiple perspectives. On the other hand, Bastani et al. (2024), also using an experimental design but focused on generative rather than rule-based AI, found that unrestricted access to generative AI can reduce independent problem-solving performance. These contrasting results are not necessarily contradictory so much as they reflect differences in the type of AI examined (adaptive/rule-based ITS versus open-ended generative AI), the outcome measured (short-term performance versus retained skill), and the instructional conditions under which AI was used supervised and scaffolded versus unrestricted access. Many of the underlying studies also rely on short-duration experimental designs or self-reported measures, which limits the extent to which their findings can be generalized to long-term learning trajectories. These methodological and conceptual inconsistencies are precisely why a systematic, rather than narrative, synthesis is needed: only a structured review can reveal whether the discrepancy in findings reflects genuine contextual moderators or merely accumulated measurement noise across the field.

More recently, researchers have highlighted the phenomenon of cognitive offloading, in which individuals increasingly delegate cognitive processes to technological systems, leading to a decline in active reasoning and analytical engagement (Bastani et al., 2024). These divergent findings suggest that the influence of AI on the development of mathematical literacy cannot be characterized simply as either entirely beneficial or entirely detrimental, but is instead moderated by how, and under what pedagogical conditions, AI is used.

What distinguishes the present review from earlier work is threefold. First, unlike prior reviews that address AI in education broadly (e.g., Zawacki-Richter et al., 2019), this review focuses specifically on mathematical literacy as the outcome construct rather than general learning effectiveness. Second, it centers on Generative AI, whose capacity for open-ended, non-deterministic output raises verification and evaluation demands on learners that earlier rule-based ITS research did not have to address. Third, rather than only cataloguing reported benefits, challenges, and future directions, this review synthesizes these threads into a candidate construct Mathematical AI Literacy intended to extend, rather than replace, the PISA definition of mathematical literacy for the AI era. Absent this explicit positioning, the contribution of the review would risk appearing incremental relative to existing educational-technology reviews.

Based on these considerations, this study aims to conduct a systematic review of existing research on the application of Artificial Intelligence in mathematics education to examine its influence on the development of mathematical literacy. Furthermore, this review seeks to identify the benefits, challenges, emerging research directions, and pedagogical implications of AI integration in mathematics learning. The findings are expected to provide a scientific foundation for developing instructional strategies that position AI as an effective learning partner while preserving students' critical thinking, mathematical reasoning, and independent learning skills in the future. To guide this synthesis, the review addresses four research questions: RQ1 What benefits of AI use are reported to influence the development of mathematical literacy? RQ2 What risks or challenges associated with AI use are reported to affect mathematical reasoning and learning autonomy? RQ3 How does the reviewed literature suggest that the construct of mathematical literacy is being reshaped in the era of Generative AI? RQ4 What research gaps remain regarding the relationship between AI and mathematical literacy that warrant further investigation? These four questions correspond directly to the four subsections of the Results and Discussion “AI as a Catalyst for Mathematical Literacy” (RQ1), “AI and the Risk of Learning Dependency” (RQ2), “Mathematical Literacy in the Age of Artificial Intelligence” (RQ3), and “Research Gaps” (RQ4) so that readers can directly evaluate whether each question is answered.

METHODS

This study employed a Systematic Literature Review (SLR) following the PRISMA 2020 guidelines (Page et al., 2021). This approach was selected because it enables a transparent, reproducible, and less biased synthesis of a large body of relevant literature. Unlike conventional narrative reviews, which often rely on authors' subjective judgment in selecting and interpreting studies, a systematic literature review requires explicit and rigorous procedures throughout every stage of the review process, from literature identification to data synthesis. The literature was retrieved from four electronic databases, namely Google Scholar, ERIC, Scopus, and the Garuda Portal. Studies were included if they were published between 2015 and 2024, written in English or Indonesian, focused on the application of Artificial Intelligence in mathematics education, available in full-text format, and published in peer-reviewed journals or official institutional reports. Conversely, studies were excluded if they did not address mathematical literacy, lacked a clearly described research methodology, or focused exclusively on educational technologies unrelated to Artificial Intelligence.

The search strategy combined keyword terms using explicit Boolean operators to balance sensitivity and specificity. In Google Scholar and ERIC, the string (“mathematical literacy” OR “mathematics literacy”) AND (“artificial intelligence” OR “AI” OR “ChatGPT” OR “generative AI”) was applied, supplemented by (“artificial intelligence” OR “AI”) AND (“mathematics education” OR “mathematics learning”). In Scopus, the equivalent field-restricted string TITLE-ABS-KEY((“mathematical literacy” OR “mathematics literacy”) AND (“artificial intelligence” OR “AI” OR “ChatGPT”)) was used. Because the inclusion criteria permitted studies in either English or Indonesian, the search conducted in the Garuda Portal additionally used Indonesian-language equivalents, namely “literasi matematika AND kecerdasan buatan,” “AI dalam pembelajaran matematika,” “ChatGPT dalam pembelajaran matematika,” “kecerdasan buatan literasi matematika,” and “kecerdasan buatan dalam pembelajaran matematika,” ensuring that Indonesian-language studies indexed in Garuda were not systematically missed.

To assess the methodological quality of the included studies, this review applied the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018), which provides screening questions and category-specific quality criteria appropriate for quantitative randomized, quantitative non-randomized, qualitative, and mixed-methods designs. Each of the 32 included studies was appraised against the criteria corresponding to its design category; studies that did not meet a minimum quality threshold were flagged in the extraction record but retained for narrative thematic synthesis, given the substantial heterogeneity of designs typical of AI-in-education research, rather than being excluded outright.

Data extraction was conducted using a structured form capturing, for each included study, the author(s), publication year, country or research context, research method, sample characteristics, the specific AI technology examined, and the main findings relevant to mathematical literacy.

Table 1. Article Selection Process

Stage	Number of Articles
Identification	287
Screening	241
Eligibility	46
Included	32

Beyond the raw counts shown in Table 1, PRISMA 2020 reporting standards require that the reason for exclusion be specified at each stage. At the Screening stage, the [n] records excluded from the initial 287 were removed primarily for reasons such as duplication across databases, titles/abstracts unrelated to Artificial Intelligence, and publications outside the 2015–2024 window. At the Eligibility stage, 14 of the 46 full-text articles assessed were excluded, primarily because they did not address mathematical literacy specifically, lacked a clearly described research methodology, or were not available in full-text format. Figure 1 presents the complete PRISMA flow diagram corresponding to these four stages.

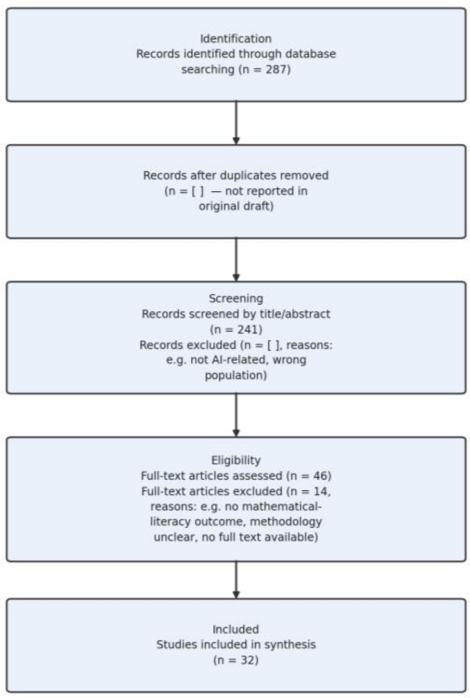

Figure 1. PRISMA 2020 Flow Diagram of the Study Selection Process

The selected studies were analyzed using thematic analysis (Braun & Clarke, 2006), which involved four sequential stages: coding, categorization, data synthesis, and interpretation of findings. This analytical approach facilitated the identification of recurring themes, research trends, reported benefits and challenges, as well as emerging directions regarding the application of Artificial Intelligence in mathematics education and its implications for the development of mathematical literacy. Figure 2 illustrates the progression of this analytical procedure from initial coding to final interpretation.

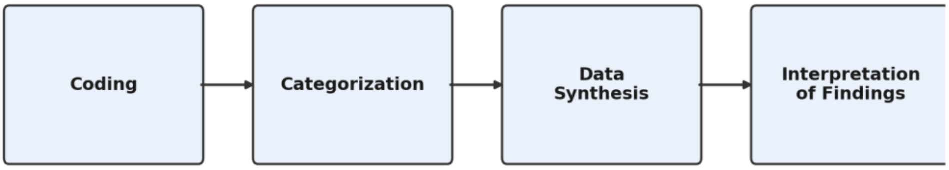

Figure 2. Stages of Thematic Analysis, from Coding to Interpretation of Findings

RESULT AND DISCUSSION

The findings presented in this section are derived from the synthesis of the selected studies included in this systematic literature review. The findings presented in this section are derived from the synthesis of the 32 studies included in this systematic literature review.

Table 2. Distribution of Research Methods Across the 32 Included Studies

Method	Number of Studies (n)	Percentage
Meta-Analysis	5	15,625 %
Experimental	16	50 %
Survey	5	15,625 %
Qualitative / Case Study	4	12,5 %
Mixed-Methods	2	6,25 %
Total	32	100%

Table 3. Summary of Illustrative Studies

No	Authors	Year	Method	Main Findings
1.	Ma et al.	2014	Meta-Analysis	Intelligent Tutoring Systems (ITS) significantly improve students' mathematics achievement.
2.	Wardat et al.	2023	Experimental	ChatGPT enhances students' conceptual understanding of mathematics.
3.	Bastani et al.	2024	Experimental	Unrestricted use of generative AI reduces students' independent problem-solving performance.

No	Authors	Year	Method	Main Findings
4.	Yeh et al.	2021	Survey	Gamified AI-supported environments increase achievement and interest in mathematics learning.
5.	Yıldız & Çakıroğlu	2022	Experimental	AI-assisted adaptive learning reduces students' mathematics anxiety.

Table 3 presents five illustrative studies selected to represent the principal methodological types identified among the 32 included studies meta-analytic (Ma et al., 2014), experimental (Wardat et al., 2023; Bastani et al., 2024; Yıldız & Çakıroğlu, 2022), and survey-based (Yeh et al., 2021) designs rather than being a random or exhaustive sample.

AI as a Catalyst for Mathematical Literacy (RQ1)

The synthesis of the selected studies indicates that a majority of the reviewed studies [X of 32] report that Artificial Intelligence contributes to the quality of mathematics education, although the mechanisms differ depending on the type of AI examined. Studies focused on rule-based, adaptive systems and studies focused on generative, conversational AI report converging benefits through different pathways, and distinguishing between them clarifies what is actually driving the reported gains.

Among rule-based systems, the meta-analysis conducted by Ma et al. (2014), synthesizing evidence across a large number of primary studies, demonstrated that Intelligent Tutoring Systems (ITS) significantly improve mathematics learning outcomes by identifying students' misconceptions, providing step-by-step guidance, and adapting problem difficulty based on learners' progress. Because this evidence is aggregated across many controlled studies, it offers comparatively strong, statistically-grounded support for the effectiveness of tightly scaffolded, algorithmically constrained AI unlike most other claims in this review, which are drawn from qualitative or single-study evidence and are described below in non-statistical terms. This pattern is also consistent with broader observations that student learning rates follow a surprisingly regular trajectory across diverse instructional contexts (Koedinger, Carvalho, Liu, & McLaughlin, 2023), suggesting that adaptive systems may be capitalizing on a structural regularity in how students learn rather than producing an idiosyncratic effect specific to any single tool. VanLehn (2011) similarly found that well-designed ITS can approach, though not always match, the effectiveness of individualized human tutoring, further situating the ITS evidence base within a longer line of tutoring-effectiveness research.

By contrast, the benefits reported for generative AI rest on a much thinner evidentiary base. Wardat et al. (2023), using a single experimental study, found that ChatGPT facilitates students' understanding of abstract mathematical concepts by providing simplified explanations, step-by-step solutions, and multiple alternative problem-solving strategies. Similarly, Yeh et al. (2021) reported that gamified, AI-supported learning environments increase both achievement and interest in mathematics, consistent with broader evidence that structured, technology-based practice can increase measurable achievement (Roschelle, Feng, Murphy, & Mason, 2016). These studies point in the same direction as the ITS evidence, but because they are individual experimental or survey studies rather than aggregated syntheses, and because generative AI's open-ended output is qualitatively different from an ITS's constrained response space, the strength of the claim that generative AI reliably improves conceptual understanding should be treated as considerably more provisional than the claim for ITS. Practical frameworks for structuring generative-AI-supported tasks (Mollick & Mollick, 2023) further suggest that the size of this benefit depends heavily on how instructors design the task, not on the technology alone.

Taken together, these findings suggest that AI possesses considerable potential to foster mathematical literacy, particularly by enhancing conceptual understanding, mathematical communication, and students' ability to connect mathematical concepts with authentic real-world contexts but that this potential is better evidenced for constrained, adaptive AI systems than for open-ended generative AI, where the same instructional benefit appears more contingent on how the tool is used.

AI and the Risk of Learning Dependency (RQ2)

Despite its numerous educational benefits, [Y of 32] of the reviewed studies also highlight potential drawbacks associated with excessive AI utilization, although these risks are not uniform across contexts and should not be generalized as an inevitable outcome of AI use. Bastani et al. (2024) reported that, under conditions of unrestricted access, students who relied heavily on generative AI exhibited weaker independent problem-solving abilities because substantial portions of the reasoning process were delegated to the technological system. This finding, however, is drawn from a specific experimental condition and does not by itself establish that AI use universally undermines reasoning; it indicates instead that the risk is concentrated in unscaffolded, unsupervised usage patterns. Even Intelligent Tutoring Systems, despite the comparatively strong evidence of their benefits discussed above, are not without limitations: Ritter et al. (2016) documented cases in which ITS struggled to handle non-routine problem types and to diagnose certain student misconceptions accurately, underscoring that the benefits summarized in Table 2a are conditional rather than unconditional. Students' own perceptions of generative AI are similarly mixed: Chan and Hu (2023) found that students value the accessibility and immediacy of generative AI while simultaneously expressing concern about their own overreliance on it, suggesting that the risk of dependency is not merely an outside observation but one that learners themselves recognize.

This phenomenon can be interpreted through the lens of cognitive load theory (Sweller, 1988), which distinguishes between intrinsic, extraneous, and germane cognitive load. When AI absorbs the germane load associated with actively constructing a solution the load responsible for schema formation it may remove precisely the cognitive effort through which mathematical reasoning develops. This offers a theoretical explanation for cognitive offloading beyond simply describing the phenomenon: the risk is not AI assistance per se, but assistance that substitutes for, rather than scaffolds, the schema-building step of learning.

Moreover, several studies emphasize that AI-generated responses are not always accurate. Generative AI models remain susceptible to producing misleading or incorrect information, commonly referred to as AI hallucinations, particularly when solving mathematical problems requiring formal proofs or complex logical reasoning. Therefore, the ability to critically verify, evaluate, and validate AI-generated responses has become an essential competency in contemporary mathematics education.

These findings indicate that the effectiveness of AI depends not only on technological sophistication but also on appropriate instructional design and the specific cognitive load conditions under which it is used. AI should function as a cognitive support tool that strengthens students' reasoning processes rather than replacing human thinking, and the reviewed evidence supports this as a context-dependent claim rather than a universal law of AI-assisted learning.

Mathematical Literacy in the Age of Artificial Intelligence (RQ3)

The findings further suggest that the emergence of AI has transformed the traditional understanding of mathematical literacy. Whereas mathematical literacy has traditionally emphasized the ability to formulate, employ, and interpret mathematics across various contexts (OECD, 2023), the AI era requires this competence to be expanded to include critical interaction with intelligent technologies.

The reviewed literature suggests that these competencies the ability to evaluate the accuracy of AI-generated responses, identify algorithmic errors, distinguish relevant from irrelevant information, use AI ethically and responsibly, and integrate AI-generated outputs into mathematical problem-solving may collectively represent an emerging construct that this review tentatively refers to as Mathematical AI Literacy. This wording reflects the nature of the evidence synthesized here: an evidence-based inference from converging patterns across the reviewed studies rather than a validated theoretical framework. These competencies demonstrate that students are expected not only to operate AI technologies effectively but also to regulate and critically evaluate their use through sound mathematical reasoning.

Accordingly, mathematical literacy in the AI era extends beyond computational proficiency and problem-solving skills. It also encompasses reflective thinking, evidence-based decision-making, and the capacity to critically validate information generated by AI systems.

Research Gaps (RQ4)

Before turning to the identified gaps, it is important to substantiate the claim that Indonesia-focused evidence is limited. Among the included studies, only a small subset explicitly situates its findings within the Indonesian context: Prahmana et al. (2023) surveyed AI use among Indonesian secondary school students, and Sumintono (2022) examined the broader digital divide affecting the equity of mathematics education in Indonesia.

Although research on AI in mathematics education has grown rapidly, this review identifies several research gaps that warrant further investigation, each arising from a different underlying limitation in the existing evidence base. First, most existing studies primarily focus on the effectiveness of AI in improving general learning outcomes, whereas relatively few directly examine its influence on the development of mathematical literacy specifically; this gap reflects a conceptual ambiguity in how prior research has operationalized “learning outcomes” rather than a lack of studies as such. Second, empirical research conducted in Indonesia remains predominantly descriptive, resulting in limited experimental evidence regarding the impact of AI on students' mathematical literacy a gap driven by insufficient empirical (particularly experimental) evidence at the national level, in contrast to a comparatively larger, though still contested, experimental literature internationally. Third, to date no standardized instrument has been established to assess Mathematical AI Literacy; researchers continue to employ diverse indicators and measurement frameworks, which stems from methodological inconsistency across studies and makes cross-study comparison difficult. Fourth, research investigating the integration of AI within the implementation of the Merdeka Curriculum, particularly concerning AI-supported mathematical literacy assessment, remains scarce, reflecting a geographical and policy-specific imbalance in the literature: most reviewed evidence originates from contexts outside Indonesia's national curriculum framework.

These four gaps differ in scope. The first and third are global in nature they concern how the field as a whole defines and measures the construct whereas the second and fourth are local to the Indonesian context and its curriculum policy. Read together, they point to a shared theoretical contribution of this review: the field currently lacks a coherent, measurable definition of what AI-era mathematical literacy actually consists of, and until such a definition is operationalized and validated, both the global debate over AI's effects and Indonesia-specific curriculum integration will remain difficult to resolve empirically. This positions the Mathematical AI Literacy construct proposed above not merely as descriptive vocabulary, but as a necessary precursor to closing these gaps.

Overall, the synthesis demonstrates that Artificial Intelligence possesses substantial potential to enhance mathematics education through personalized instruction, improved conceptual understanding, and adaptive feedback mechanisms. Nevertheless, its implementation should be accompanied by pedagogical strategies that actively promote critical thinking and mathematical reasoning. In this regard, AI should be positioned as a learning partner that supports students' cognitive development rather than as a substitute for human reasoning.

CONCLUSION

This systematic literature review of 32 studies indicates that Artificial Intelligence makes a meaningful, though unevenly evidenced, contribution to the development of mathematical literacy. The strongest support comes from aggregated evidence on constrained, adaptive systems such as Intelligent Tutoring Systems, while evidence for generative AI tools such as ChatGPT remains more provisional and contingent on how these tools are used. Beyond confirming that AI can enhance conceptual understanding, timely feedback, and problem-solving support when integrated with appropriate pedagogical strategies, the principal contribution of this review is conceptual: it proposes that mathematical literacy in the AI era must be understood as extending beyond the traditional PISA formulation to include the ability to evaluate the accuracy of AI-generated responses, identify the limitations and potential errors of AI systems, use AI ethically and responsibly, and critically integrate AI-generated information into mathematical problem-solving. The reviewed evidence suggests, though does not yet definitively establish, that these competencies may collectively constitute an emerging construct Mathematical AI Literacy whose theoretical boundaries and empirical measurement remain to be validated in future work.

At the same time, the review shows that excessive or unscaffolded AI use may encourage cognitive offloading, reduce learning autonomy, and weaken independent mathematical reasoning, a risk best explained through cognitive load theory rather than treated as an inherent property of AI itself. This reinforces the conclusion that AI should not replace students' cognitive processes in mathematics learning, but should instead serve as a supportive tool that facilitates exploration, reflection, and verification throughout the problem-solving process.

The gaps identified in this review the conceptual ambiguity in how mathematical-literacy outcomes are operationalized, the limited experimental evidence from Indonesia, the absence of a standardized Mathematical AI Literacy instrument, and the scarcity of research on AI integration within the Merdeka Curriculum collectively point toward a single overarching priority for future research: the development and validation of a measurable Mathematical AI Literacy framework. Future studies are therefore recommended to develop valid and reliable instruments for assessing this construct, conduct experimental and mixed-methods research across educational levels and, specifically within Indonesia, in relation to national curriculum policy, and design instructional models that integrate AI utilization with the continuous development of critical thinking, mathematical reasoning, and mathematical literacy.

REFERENCES

Anwari, A. M., & Nuriadin, I. (2025). Kajian Literatur tentang Integrasi AI dalam Pendidikan Matematika: Potensi dan Tantangan. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *10*(2), 245–256. https://doi.org/10.30998/jkpm.v10i2.27633

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabir, O., & Marber, R. (2024). *Generative AI can harm learning*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4895486

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *Computers and Education: Artificial Intelligence, 4*, 100134. https://doi.org/10.1016/j.caeai.2023.100134

English, L. D. (2023). Ways of thinking in STEM-based problem solving. *ZDM–Mathematics Education, 55*(7), 1219–1230. https://doi.org/10.1007/s11858-023-01474-7

Febrianti, K. R., Azizah, N., & Rusadi, F. (2025). Pemanfaatan Kecerdasan Buatan Artificial Intelligence (AI) dalam Membantu Kinerja Pembelajaran. *Jurnal Inovasi Pendidikan* *dan Teknologi Informasi (JIPTI)*, *6*(1), 210–226. https://doi.org/10.52060/jipti.v6i1.2883

Gadanidis, G. (2017). Artificial intelligence as a mathematics education platform. *For the Learning of Mathematics, 37*(3), 16–23.

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information, 34*(4), 285–291. https://doi.org/10.3233/EFI-180221

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., Pfeiffer, F., 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*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences, 120*(13), e2221311120. https://doi.org/10.1073/pnas.2221311120

Lesh, R., Hamilton, E., & Kaput, J. (2020). *Foundations for the future in mathematics education*. Routledge. https://doi.org/10.4324/9781003064497

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. https://doi.org/10.1037/a0037123

Mollick, E. R., & Mollick, L. (2023). *Assigning AI: Seven approaches for students, with prompts*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4475995

OECD. (2023). *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

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

Prahmana, R. C. I., Istiandaru, A., Syarifah, L. L., & Hamdi, S. (2023). Penggunaan kecerdasan buatan dalam pembelajaran matematika: Survei nasional pada siswa sekolah menengah di Indonesia. *Jurnal Riset Pendidikan Matematika, 10*(2), 145–162.

Rayhan, N. C., & Juandi, D. (2023). Students learning obstacles related to mathematical literacy problem: A systematic literature review. *Mathline: Jurnal Matematika Dan Pendidikan Matematika*, *8*(2), 457–472.

Ritter, S., Fancsali, S. E., Yudelson, M. V., Cayton-Hodges, G., Joshi, A., & Berman, S. R. (2016). The trouble with math: A practical demonstration of the limitations of AI tutoring systems. In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 456–461).

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open, 2*(4), 1–12. https://doi.org/10.1177/2332858416673968

Sumintono, B. (2022). Kesenjangan digital dan pemerataan kualitas pendidikan matematika di Indonesia: Tantangan era pasca-pandemi. *Jurnal Pendidikan dan Kebudayaan, 7*(1), 1–18.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

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

Yeh, C. Y. C., Cheng, H. N. H., Chen, Z. H., Liao, C. C. Y., & Chan, T. W. (2021). Enhancing achievement and interest in mathematics learning through Math-Island. *Research and Practice in Technology Enhanced Learning, 16*(1), 5. https://doi.org/10.1186/s41039-019-0100-9

Yıldız, B., & Çakıroğlu, Ü. (2022). Effects of AI-based adaptive learning on math anxiety and motivation in middle school students. *Computers & Education, 178*, 104393. https://doi.org/10.1016/j.compedu.2021.104393

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0