Journal of Authentic Research
https://journal-center.litpam.com/index.php/jar/index

August 2026 Vol. 5, No. 3
e-ISSN: 2828-3724
pp. 4930-4944

Artificial Intelligence Integration in STEM Learning: Pedagogical Practices, Learning Outcomes, and Challenges

[1]I Ketut Sukarma, [2*]Hulyadi

[1]Master's Program in Science Education, Universitas Pendidikan Mandalika
[2]Chemistry Education Study Program, Mandalika University of Education
*Corresponding Author e-mail: hulyadi@undikma.ac.id

Received: June 2026; Revised: July 2026; Published: August 2026

Abstract

Advances in artificial intelligence (AI) are expanding how Science, Technology, Engineering, and Mathematics (STEM) learning is designed, enacted, and assessed. This review synthesizes AI applications, pedagogical practices, learning outcomes, and implementation challenges in STEM education. A structured integrative literature review was conducted on 30 core publications from 2019–2025. Literature was identified through scholarly metadata services, education indexes, and publisher websites using combinations of STEM education, artificial intelligence, generative AI, intelligent tutoring system, adaptive learning, learning analytics, educational robotics, and learning outcomes. Publications were included when they examined AI use or AI learning in science, mathematics, engineering, technology, or integrated STEM contexts. The thematic synthesis identified four dominant application patterns: intelligent tutoring and adaptive learning, generative AI for explanation and code production, AI-supported analytics and assessment, and robotics or AI literacy activities. AI can accelerate feedback, support differentiation, broaden idea exploration, and improve task completion. Evidence for durable conceptual understanding, transfer, creativity, and long-term problem-solving remains uneven. Benefits are most credible when AI functions as scaffolding within inquiry, problem-based learning, project-based learning, or engineering design and when learners must verify sources, justify decisions, and reflect on AI contributions. Major challenges include over-reliance, hallucination, bias, privacy, academic integrity, unequal access, and limited teacher capacity. AI should therefore be treated as a human-supervised cognitive partner rather than a substitute for learners' reasoning.

Keywords: STEM education; artificial intelligence; generative AI; pedagogical practices; learning outcomes

How to Cite: Sukarma, I. K. ., & Hulyadi, H. Integrasi Artificial Intelligence dalam Pembelajaran STEM: Praktik Pedagogis, Hasil Belajar, dan Tantangan. *Journal of Authentic Research*, 4930-4944. https://doi.org/10.36312/ydbcc937

. https://doi.org/10.36312/ydbcc937

PENDAHULUAN

Data-driven social and economic change requires students not only to master concepts but also to formulate problems, model phenomena, evaluate evidence, design solutions, and use technology responsibly. The STEM approach provides an interdisciplinary framework for connecting science and mathematics concepts with engineering practices and the use of technology in authentic problems. At the same time, artificial intelligence (AI) has evolved from intelligent tutoring systems and adaptive learning toward predictive analytics, computer vision, robotics, and generative models capable of producing text, images, code, and explanations. The literature on AI in education indicates a shift from paradigms in which AI directs students, to AI that supports collaboration, and ultimately to AI that empowers students as the primary decision makers (Hwang et al., 2020; Ouyang & Jiao, 2021). In STEM learning, this shift is important because learning activities should ideally extend beyond the final answer. Students need to construct models, test hypotheses, analyze data, evaluate design failures, and justify their solutions. Thus, the educational value of AI is determined not by how quickly it generates answers, but by its capacity to strengthen the epistemic processes and engineering practices at the core of STEM learning.

Despite this substantial potential, the adoption of AI in education often advances more rapidly than pedagogical design and evidence of learning. Previous reviews have found that many studies of AI in education focus on technological development, performance prediction, and personalization, while paying less attention to teacher perspectives, learning processes, and connections with learning theory (Zawacki-Richter et al., 2019; Chen et al., 2020; Zhai et al., 2021). The emergence of generative AI adds further complexity. Language models can help explain concepts, generate examples, write code, or propose design alternatives, but they can also produce inaccurate information and fabricated references (Day, 2023; Kasneci et al., 2023). In STEM classrooms, seemingly convincing outputs may conceal unit errors, invalid physical assumptions, fragile algorithms, or unsafe experimental procedures. Another risk arises when students submit entire problems to AI and accept the solutions without reconstructing the reasoning. This pattern accelerates task completion but may reduce opportunities to experience productive struggle, engage in debugging, and build resilience in problem-solving. Therefore, the central question is not whether AI is used, but how AI functions, task structures, teacher scaffolding, and assessment are designed so that AI use continues to generate meaningful cognitive activity.

Research on STEM and AI has developed along several streams that are not yet fully connected. Some literature examines AI as content to be learned through AI literacy, programming, robotics, and introductions to machine learning (Touretzky et al., 2019; Williams et al., 2019; Long & Magerko, 2020; Ng et al., 2021; Su et al., 2022). Another stream positions AI as a tool for personalization, analytics, intelligent tutoring, assessment, or content generation (Crompton & Burke, 2023; Chiu et al., 2023). Reviews of generative AI tend to address education in general, while disciplinary studies are dispersed across science, mathematics, engineering, and programming education (Cooper, 2023; Qadir, 2023; Wardat et al., 2023). Consequently, there is still no adequate mapping of the relationships among AI types, their pedagogical positions, STEM learning models, and the types of learning outcomes that can be supported by evidence. The novelty of this review lies in its synthesis across AI functions and STEM disciplines through a pedagogical lens. AI is analyzed not merely as a technological product, but as a tool, tutor, cognitive partner, learning object, and assessment infrastructure. This lens makes it possible to distinguish immediate improvements in task performance from the development of deeper competencies, such as transfer, creativity, data literacy, evidence evaluation, metacognition, and responsible AI use.

Based on these gaps, this review aims to: (1) map the forms of AI application in STEM learning; (2) analyze pedagogical practices that integrate AI with inquiry, problem-based learning, project-based learning, engineering design, and personalized learning; (3) synthesize reported effects on knowledge, skills, attitudes, and AI literacy; (4) identify pedagogical, technical, social, and ethical risks; and (5) formulate a conceptual framework and research agenda. The review questions are as follows: How is AI positioned in STEM learning activities? Which learning models are most consistent with meaningful AI use? Which learning outcomes are supported by evidence, and which still require further research? What challenges must be addressed at the levels of teachers, schools, curricula, and policy? The scope covers primary, secondary, and higher education in science, mathematics, engineering, technology, programming, and integrated STEM. The review does not assess the technical performance of AI models independently of their educational contexts. Its primary focus is the relationship among technological affordances, activity design, human roles, and the quality of learning outcomes.

METHODS

Review Design and Analytical Questions

This study employed a structured integrative literature review. This design was selected because the evidence on AI in STEM learning includes empirical research, design studies, system evaluations, conceptual papers, and systematic reviews. An integrative review enables different types of evidence to be compared according to pedagogical functions and learning outcomes without treating methodologically distinct designs as equivalent. The units of analysis were scholarly publications that discussed AI as a learning tool, tutoring system, analytics or assessment system, generative partner, or AI literacy content within STEM contexts. The analysis focused on six categories: contextual characteristics; type of AI; the position of AI in learning activities; pedagogical model; learning outcomes; and implementation challenges. A meta-analysis was not conducted because the selected publications were highly heterogeneous in effect sizes, instruments, intervention durations, and units of analysis.

Literature Search Strategy

The search covered publications from 2019 to 2025 through scholarly metadata services, education indexes, and publisher websites. This period was selected to capture developments in AI in education before and after the widespread adoption of generative AI. The primary search string combined three groups of terms: ("STEM education" OR "science education" OR "mathematics education" OR "engineering education" OR "computing education") AND ("artificial intelligence" OR "generative AI" OR "ChatGPT" OR "intelligent tutoring system" OR "adaptive learning" OR "learning analytics" OR "educational robotics") AND ("learning" OR "teaching" OR "assessment" OR "learning outcomes"). Additional searches used specific technologies and learning outcomes, including programming, conceptual understanding, problem-solving, computational thinking, AI literacy, feedback, and academic integrity. Reference lists from the principal reviews were examined backward to identify relevant foundational studies. Non-peer-reviewed sources were used only for authoritative policy guidance, particularly UNESCO publications.

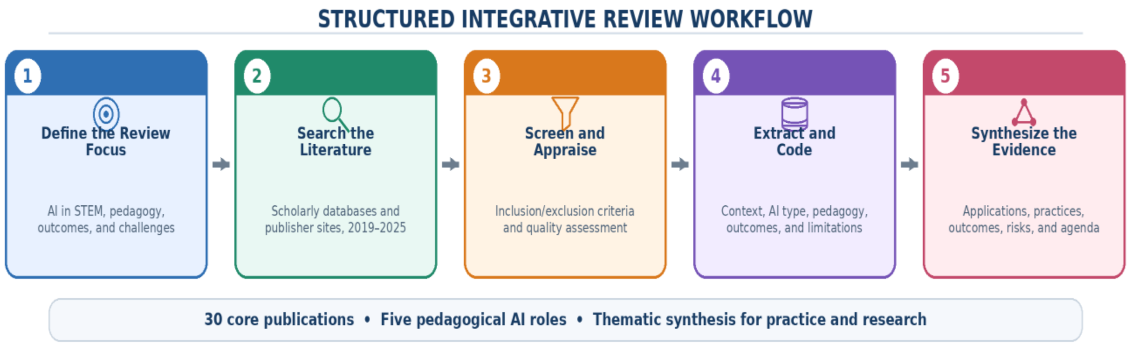

Figure 1. Literature search and synthesis process

Selection Criteria and Quality Appraisal

Publications were selected when they met all of the following criteria: they addressed the use or teaching of AI in a STEM context; provided a description of methods or a conceptual argument that could be evaluated; were published in journals, scholarly proceedings, or authoritative policy reports; were available in English or Indonesian; and included verifiable bibliographic information. Publications were excluded when they discussed only algorithmic performance without a learning context, consisted of popular opinion, did not explain their methods, or duplicated the same report. Thirty core publications were used in the main synthesis. Foundational studies published before the primary period were used sparingly to explain the effectiveness of intelligent tutoring systems and the development of AI curricula. Quality was appraised according to the clarity of objectives, appropriateness of methods, adequacy of participant and context descriptions, transparency of instruments or data, alignment between findings and evidence, and acknowledgment of limitations. Studies with claims that exceeded their data were not used as the basis for causal conclusions.

Table 1. Publication inclusion and exclusion criteria

Aspect	Inclusion criteria	Exclusion criteria
Topic	AI is used or taught in science, mathematics, engineering, technology, programming, or integrated STEM.	AI is discussed without an educational context or without a connection to STEM.
Source type	Journal articles, scholarly proceedings, systematic reviews, and authoritative policy guidelines.	News reports, opinions, product promotions, and manuscripts without verifiable methods or sources.
Period	Publications from 2019–2025; older sources only for essential foundations.	Older sources that provide no conceptual or methodological contribution.
Data sought	Type of AI, pedagogical design, context, participants, learning outcomes, challenges, and implications.	Technology descriptions without learning activities, participants, or consequences.
Language and access	English- or Indonesian-language publications with adequate abstracts/full text and metadata.	Incomplete metadata or content that cannot be evaluated.

Data Extraction and Synthesis

Each publication was coded by year, country or region, educational level, STEM discipline, type of AI, learning model, form of human-AI interaction, outcome variables, findings, and limitations. The coding distinguished five pedagogical positions of AI: productivity tool, tutor or scaffolding provider, cognitive partner, assessment/analytics system, and AI literacy object. The results were then synthesized thematically through constant comparison. Recurring findings were grouped as benefits, conditions for success, and risks. Particular attention was paid to the distinction between immediate task performance and learning demonstrated through retention, transfer, conceptual explanation, or the solution of novel problems. The final synthesis generated themes related to applications, pedagogical practices, learning outcomes, changes in teacher and student roles, ethical and technical challenges, and the research agenda.

RESULTS AND DISCUSSION

Characteristics and Mapping of the Evidence

The literature shows a rapid shift in focus. Publications before 2022 frequently addressed intelligent tutoring systems, prediction, personalization, robotics, and AI literacy. After 2022, attention shifted toward generative models and their implications for conceptual explanation, programming, scientific writing, assessment, and academic integrity. Although the term AI is used broadly, the technologies studied are not always equivalent. Tutoring systems generally operate within structured domains and solution steps; learning analytics processes activity traces; robotics connects programming, sensors, and design; whereas generative models are open-ended and produce probabilistic responses. These differences affect the type of scaffolding and the validity of outcomes. General reviews indicate that AI can support personalization, feedback, and decision-making, while also emphasizing the scarcity of studies linking technology with learning theory and teachers' needs (Hwang et al., 2020; Crompton & Burke, 2023; Chiu et al., 2023).

Within STEM contexts, the most direct evidence is found in structured mathematics and programming because answers or code can be tested relatively quickly. Intelligent tutoring systems have a more mature evidence base than generative AI; earlier meta-analyses reported positive effects of ITS, although these effects depend on the alignment of the system, curriculum, and outcome measures (Kulik & Fletcher, 2016). For generative AI, early studies indicate benefits for task completion and the production of learning resources, but they do not always demonstrate conceptual understanding or transfer. A programming study involving Codex found that AI could improve success in code writing without a clear decline in manual modification tasks; however, long-term retention and patterns of dependence still require examination (Kazemitabaar et al., 2023). These findings reinforce the need to distinguish product effectiveness from the quality of the learning process.

Table 2. Mapping AI functions, STEM activities, and the quality of evidence

AI function	STEM activities	Commonly reported benefits	Evidence notes
Intelligent/adaptive tutoring	Step-by-step practice, hints, personalized learning paths, and error diagnosis.	Rapid feedback, mastery-aligned practice, and improved domain performance.	The evidence base is relatively mature; outcomes depend on curricular integration and feedback quality.
Generative AI/LLMs	Explanations, brainstorming, question generation, code, hypotheses, and reports.	Efficiency, access to explanations, exploration of alternatives, and debugging support.	Long-term evidence is limited; risks of incorrect answers and cognitive offloading are high.
Learning analytics	Detecting difficulties, monitoring engagement, and predicting risk.	Information for teacher intervention and student reflection.	Prediction does not automatically improve learning; pedagogical action and transparency are required.
Automated assessment	Scoring answers, code, and reports, and providing feedback.	Scalability and rapid feedback.	Validity, bias, rubric clarity, and appeal mechanisms must be ensured.
Robotics/computer vision	Sensor programming, robot design, and object and motion recognition.	Connections among concepts, computational thinking, collaboration, and engineering design.	Requires devices, time, technical support, and competent teachers.
AI as learning content	Training simple models, auditing bias, comparing outputs, and examining data ethics.	AI literacy, systems understanding, critical evaluation, and digital responsibility.	Requires an age-appropriate curriculum and a balance among concepts, applications, and ethics.

AI Applications and Pedagogical Practices in STEM Learning

AI is most meaningful when its function is derived from learning objectives. In inquiry-based learning, AI can help generate initial questions, suggest variables, simulate phenomena, or organize data. Nevertheless, students must still select viable questions, control variables, and judge whether data patterns support a claim. In problem-based and project-based learning, AI can provide alternative solutions, code examples, or prototype feedback. In engineering design, AI supports iteration, but decisions about criteria, trade-offs, safety, and social impact cannot be delegated entirely to the system. The three-paradigm AIEd framework helps explain these differences: AI-directed approaches are suitable for structured practice; AI-supported approaches assist knowledge construction; and AI-empowered approaches require student agency and critical evaluation (Ouyang & Jiao, 2021). STEM learning should move toward the latter two positions, particularly for open-ended tasks.

Disciplinary evidence reveals a similar pattern. In science education, generative AI can function as a source of explanation and an object of critical evaluation, but its answers must be compared with evidence and primary sources (Cooper, 2023; Aptyka et al., 2025). In mathematics, chatbots can present solution steps and alternative examples, yet procedural accuracy does not guarantee relational understanding; students must explain their reasoning, create alternative representations, and check solutions (Wardat et al., 2023). In engineering education, AI supports ideation, programming, documentation, and simulation, but curricula must strengthen verification, ethics, and assessment design (Qadir, 2023). In programming, code models can generate useful exercises and explanations for instructors, provided that quality is reviewed before use (Sarsa et al., 2022). Pedagogically, these findings reject the use of AI as an answer machine. Stronger activities require students to predict outputs before using AI, test the outputs, identify errors, revise prompts or models, and explain their final decisions.

Table 3. Examples of key evidence in STEM contexts

Study	**Context and design**	**Relevant findings**	**Pedagogical implications**
Finnie-Ansley et al. (2022)	Evaluation of Codex's ability to solve introductory programming tasks.	The model can solve many basic tasks, making routine questions easy to automate.	Assessment should evaluate explanation, debugging, modification, and transfer.
Sarsa et al. (2022)	Generation of programming exercises and code explanations using language models.	Many outputs are plausible and useful but still require instructor review.	AI is suitable for supporting resource production, not automatic publication without curation.
Kazemitabaar et al. (2023)	Controlled experiment with 69 novices aged 10–17.	AI access improved task completion and code-writing scores; retention requires further investigation.	Use AI with modification tasks, explanations, and delayed assessment.
Cooper (2023)	Analysis of the implications of ChatGPT for science education.	AI expands access to explanations but can generate inaccuracies.	Include evidence verification and response-quality evaluation as learning objectives.
Wardat et al. (2023)	Review of ChatGPT use in mathematics learning.	Potential for examples, dialogue, and personalization is accompanied by risks of error and dependence.	Focus on reasoning, representation, and solution checking.

Study	Context and design	Relevant findings	Pedagogical implications
Qadir (2023)	Analysis of opportunities and risks in engineering education.	AI supports design, coding, and feedback but challenges assessment and integrity.	Redesign tasks to emphasize design decisions, justification, and professional responsibility.
Day (2023)	Test of references generated by ChatGPT.	References appeared relevant but were nonexistent.	All AI-generated citations must be verified through databases or publisher websites.
Nazaretsky et al. (2022)	Study of teachers' trust in AI-powered educational technology and professional development.	Knowledge and training influence calibrated trust.	Training should cover how AI works, its limitations, output interpretation, and pedagogical decisions.
Long & Magerko (2020); Ng et al. (2021)	AI literacy competency frameworks.	Literacy includes understanding, using, evaluating, creating, and considering ethics.	AI literacy should be embedded in STEM activities rather than treated as separate add-on content.
Aptyka et al. (2025)	Learning about evolution among secondary school students using ChatGPT compared with conventional search.	Conceptual gains were reported as limited, but attitudes toward learning with AI were more positive.	The added value of AI should be tested through understanding, transfer, and dialogue quality rather than tool novelty.

Effects on Learning Outcomes and Competencies

Overall, the publications report four groups of outcomes: academic performance or task completion, conceptual understanding, thinking and engineering skills, and attitudes or engagement. The most visible effects involve speed, the number of completed tasks, the initial quality of products, and access to feedback. These outcomes are valuable, particularly in large classes and for students who require additional support. However, higher performance while AI is available is not necessarily equivalent to learning. When AI generates complete steps, code, or explanations, students may reach correct answers without constructing mental models that can be applied to new situations. Rapid reviews of ChatGPT in education have identified many claims of benefit but have emphasized the limitations of empirical studies, short intervention periods, and the dominance of user perceptions (Lo, 2023; Tlili et al., 2023; Mogavi et al., 2024). STEM-AI research should therefore employ delayed post-tests, transfer tasks, think-aloud protocols, log analysis, and assessments of argument quality rather than relying only on product scores.

Effects on creativity and problem-solving are also conditional. Generative AI can broaden the idea space, present analogies, or offer multiple strategies. Positive effects arise when students compare alternatives, combine ideas, and justify their choices. Conversely, the first fluent idea may create anchoring and homogenize solutions. In programming, AI reduces the syntactic burden so that novices can focus on program goals, but it can also reduce opportunities to understand errors when code is accepted without debugging. In engineering projects, the ease of generating designs must be balanced by testing against criteria, local data, costs, safety, and sustainability. Recommended outcome indicators therefore include conceptual accuracy, the quality of questions, evidence evaluation, explainable originality, the ability to detect AI errors, transfer, metacognition, and collaboration. AI literacy is a cross-cutting outcome because students must understand the capabilities, limitations, biases, and consequences of using such systems (Long & Magerko, 2020; Ng et al., 2021).

Changing Roles of Teachers, Students, and Assessment

AI integration shifts the teacher's role from information provider to experience designer, source curator, scaffolding provider, validator, and ethical steward. This shift does not reduce the importance of teachers; rather, pedagogical decisions become more complex because teachers must determine when AI is permitted, which functions support the objectives, what data may be entered, and what evidence of learning must be collected. Teacher trust must be calibrated: complete skepticism may close off opportunities, whereas trust without understanding may lead to risky use. Professional development programs that explain how AI works, its limitations, how to interpret outputs, and examples of learning activities can improve readiness (Nazaretsky et al., 2022). The TPACK framework also needs to be expanded to include the generative and probabilistic nature of AI, so that technological knowledge is always connected with content, pedagogy, context, and ethical considerations (Mishra et al., 2023; Southworth et al., 2023).

Students need to act as problem identifiers, data collectors, claim testers, solution designers, and evaluators of AI outputs. These roles should be evidenced through process artifacts, including prompt records, design versions, test results, error annotations, verification sources, and reflections on human and AI contributions. Assessment then shifts from a single product toward triangulating process and product. Relevant strategies include oral presentations, live demonstrations, follow-up questions, tasks using local data, solution modification, design journals, and comparisons between students' initial responses and their responses after AI support. Assessment policies should distinguish among prohibited use, permitted use with disclosure, and required use as part of AI literacy objectives. This approach is more educational than relying on AI-text detectors, which cannot serve as the sole evidence. Academic integrity concerns remain real, but effective responses lie in task design, transparency, and dialogue about the learning process (Cotton et al., 2024; UNESCO, 2023).

Pedagogical, Technical, Ethical, and Equity Challenges

The primary pedagogical challenge is misalignment between AI affordances and learning objectives. AI is often used because it is available rather than because it is necessary. This produces the same tasks as traditional methods, but their answers can be automated. The risk of dependence increases when students lack sufficient prior knowledge to evaluate outputs. Hallucinations, fabricated references, and overly confident explanations are particular concerns in science and engineering because errors may affect safety or experimental validity (Day, 2023; Kasneci et al., 2023; Farrokhnia et al., 2024). Technical challenges include connectivity, devices, costs, changes in model versions, language, and integration with learning management systems. Replication is also difficult because commercial systems change and their responses may vary over time. Researchers need to document system versions, access dates, settings, prompts, and verification mechanisms.

Ethical challenges include data privacy, copyright, algorithmic bias, decision transparency, and academic integrity. Student data, laboratory reports, or sensitive information should not be entered into services without a clear legal basis and adequate safeguards. Bias may appear in the selection of examples, language, cultural contexts, and career recommendations. Ethical principles for AIEd require beneficence, nonmaleficence, justice, autonomy, transparency, and accountability (Nguyen et al., 2023). Equity is a major concern because students with better devices, connectivity, dominant-language proficiency, and access to paid services receive stronger support. Schools need to provide equivalent alternatives, data-protection policies, complaint mechanisms, and human oversight. AI should not be used for high-stakes decisions about students without professional review. UNESCO guidance emphasizes human-centered, age-appropriate approaches that preserve user agency (UNESCO, 2021, 2023).

Conceptual Framework for STEM-AI Integration

Based on the synthesis, responsible STEM-AI integration begins with authentic problems and learning objectives rather than with application selection. Students integrate STEM concepts through investigation and data. AI may then function as a simulation/analysis tool, tutor/feedback provider, or generative partner. This support is directed toward engineering design, testing, and revision. A mandatory verification stage requires students to evaluate sources, test outputs, explain their reasoning, and reflect on AI contributions. Expected outcomes include conceptual understanding, problem-solving, creativity, computational thinking, AI literacy, and collaboration. The entire process is bounded by safeguards for ethics, privacy, transparency, academic integrity, equitable access, and human oversight. The framework in Figure 2 positions AI as a pedagogical design component that can be activated or constrained according to the learning objectives.

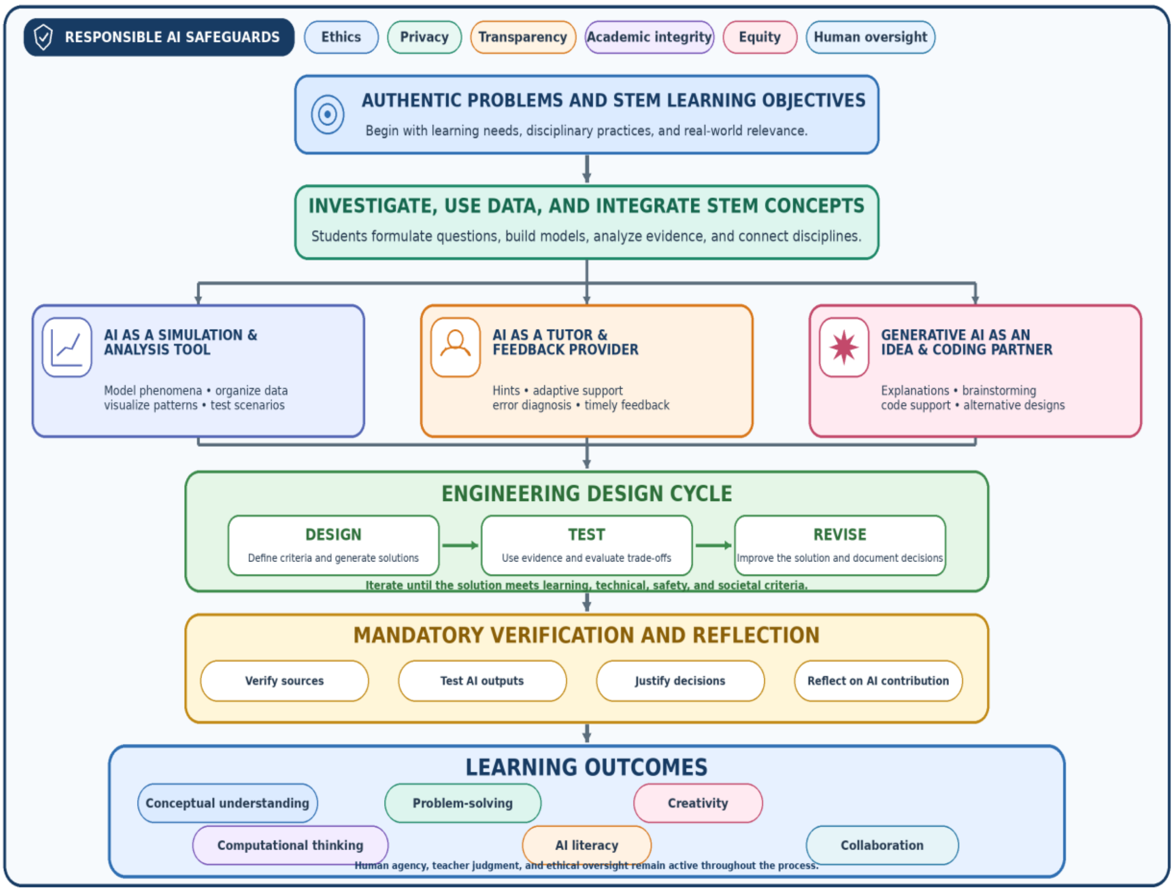

Figure 2. Conceptual framework for AI integration in STEM learning

Research Gaps and Agenda

The literature remains dominated by short-term studies, higher education contexts, immediate performance measures, and English-language technologies. Future research needs to examine retention and transfer, compare types of scaffolding, and determine when AI helps or hinders students with different levels of prior knowledge. Research designs should include no-AI, unrestricted-AI, and structured-AI conditions; interaction logs; process assessment; and interviews or think-aloud protocols. Research should be expanded in primary and secondary schools, regions with limited infrastructure, local-language settings, inclusive education, and non-Western cultural contexts. In addition to cognitive outcomes, studies should measure agency, creativity, collaboration, calibrated trust, AI literacy, and ethical behavior. System development should also involve teachers and students through co-design so that technology responds to classroom needs rather than merely reflecting model capabilities.

Table 4. Research gaps and development priorities

Current findings	Limitations	Research/development priorities
AI improves task completion and feedback speed.	Retention, transfer, and conceptual understanding remain underexamined.	Delayed post-tests, transfer tasks, and reasoning analysis.

Current findings	Limitations	Research/development priorities
Many studies are conducted in higher education and programming.	Evidence in primary/lower secondary education and integrated STEM is limited.	Studies across educational levels, disciplines, and local contexts.
Generative AI supports ideation and explanation.	Risks of anchoring, idea homogenization, and cognitive offloading.	Compare prompt scaffolding, AI critique, and metacognitive reflection.
Analytics can identify students at risk.	Predictions are not always followed by effective interventions.	Transparently test the prediction–decision–intervention–outcome chain.
Training improves teacher readiness.	Professional development models have rarely been tested longitudinally.	TPACK-AI programs integrating technology, pedagogy, content, and ethics.
Integrity policies are emerging.	Rules often focus on prohibition and detection.	Authentic assessment, disclosure of AI use, and appeal mechanisms.
AI may expand access.	Access to devices, languages, costs, and services is unequal.	Low-bandwidth design, multilingual support, and equity evaluation.

CONCLUSION

AI integration in STEM learning encompasses intelligent and adaptive tutoring systems, generative AI, learning analytics, automated assessment, robotics, computer vision, and AI literacy activities. These technologies can accelerate feedback, support differentiation, broaden exploration, and assist task completion. However, the evidence does not support the assumption that the mere presence of AI automatically improves conceptual understanding, creativity, or problem-solving. The most defensible benefits occur when AI is positioned as scaffolding within inquiry, problem-based learning, project-based learning, or engineering design, while students continue to construct explanations, test solutions, verify sources, and reflect on AI use. Teachers retain a central role as designers, curators, validators, and ethical supervisors. Successful implementation requires AI literacy, process-based assessment design, data protection, transparency, equitable access, and human oversight. The main contribution of this review is a framework that connects authentic problems, the integration of STEM concepts, AI support, design and verification processes, learning outcomes, and ethical safeguards. This framework can help teachers, curriculum developers, and researchers determine whether AI use genuinely strengthens learning or merely accelerates answer production.

RECOMMENDATIONS

1. Teachers should select AI functions according to learning objectives, require verification, and assess traces of both the process and the product.
2. Schools should establish policies on AI use, data protection, equitable access, and continuing professional development.

3. Curriculum developers should integrate AI literacy, data literacy, computational thinking, engineering design, and ethics.
4. Researchers should employ longitudinal designs, transfer tasks, log analysis, and more diverse contexts.
5. Technology developers should provide transparency, teacher control, local-language support, and audit and appeal mechanisms.

REFERENCES

Aptyka, H., Großschedl, J., & Hartelt, T. (2025). Bugbear or surefire success? Secondary school students' conceptual learning about evolution with ChatGPT. International Journal of Science Education. Advance online publication.

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

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education: Artificial Intelligence, 4, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. Journal of Science Education and Technology, 32, 444–452. https://doi.org/10.1007/s10956-023-10039-y

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

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

Day, T. (2023). A preliminary investigation of fake peer-reviewed citations and references generated by ChatGPT. The Professional Geographer, 75(6), 1024–1035. https://doi.org/10.1080/00330124.2023.2190373

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

Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., & Prather, J. (2022). The robots are coming: Exploring the implications of OpenAI Codex on introductory programming. In Proceedings of the 24th Australasian Computing Education Conference (pp. 10–19). ACM. https://doi.org/10.1145/3511861.3511863

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. Computers and Education: Artificial Intelligence, 1, 100001. https://doi.org/10.1016/j.caeai.2020.100001

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

Kazemitabaar, M., Chow, J., Ma, C. K. T., Ericson, B. J., Weintrop, D., & Grossman, T. (2023). Studying the effect of AI code generators on supporting novice learners in introductory programming. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3544548.3580919

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. Review of Educational Research, 86(1), 42–78. https://doi.org/10.3102/0034654315581420

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

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. Journal of Digital Learning in Teacher Education, 39(4), 235–251.

Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H. S., Tlili, A., Bassanelli, S., Bucchiarone, A., Gujar, S., Nacke, L. E., & Hui, P. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. Computers in Human Behavior: Artificial Humans, 2(1), 100027.

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. British Journal of Educational Technology, 53(4), 914–931.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. Education and Information Technologies, 28, 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In 2023 IEEE Global Engineering Education Conference (EDUCON) (pp. 1–9). IEEE.

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. In Proceedings of the 2022 ACM Conference on International Computing Education Research (Vol. 1, pp. 27–43). ACM. https://doi.org/10.1145/3501385.3543957

Southworth, J., Migliaccio, K., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI across the curriculum:

Transforming the higher education landscape via innovation in AI literacy. Computers and Education: Artificial Intelligence, 4, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Su, J., Zhong, Y., & Ng, D. T. K. (2022). A meta-review of literature on educational approaches for teaching AI at the K–12 levels in the Asia-Pacific region. Computers and Education: Artificial Intelligence, 3, 100065.

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

Touretzky, D. S., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K–12: What should every child know about AI? Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795

UNESCO. (2021). AI and education: Guidance for policy-makers. UNESCO.

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

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

Williams, R., Park, H. W., Oh, L., & Breazeal, C. (2019). PopBots: Designing an artificial intelligence curriculum for early childhood education. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 9729–9736. https://doi.org/10.1609/aaai.v33i01.33019729

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

Zhai, X., Chu, X., Chai, C. S., Jong, M. S.-Y., Istenic, A., Spector, J. M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. Complexity, 2021, 8812542. https://doi.org/10.1155/2021/8812542