

Editorial**Artificial Intelligence in STEM Education — Innovations, Pedagogy, Ethics, and Equity**Ali Bicer¹, Stamatios Papadakis², Tugce Aldemir¹, Ariffin Abdul Mutalib³¹ *Texas A & M University, UNITED STATES*² *University of Crete, GREECE*³ *Universiti Utara Malaysia, MALAYSIA***Corresponding Author:** alibicer@tamu.edu, stpapadakis@gmail.com, taldemir@tamu.edu, am.ariffin@uum.edu.my**This paper is part of the special issue:** [AI and STEM: Exploring the Impact of Artificial Intelligence on Science, Technology, Engineering, and Mathematics Education](#)**Citation:** Bicer, A., Papadakis, S., Aldemir, T., & Mutalib, A. A. (2026). Artificial intelligence in STEM education — Innovations, pedagogy, ethics, and equity. *European Journal of STEM Education*, 11(1), Article 38. <https://doi.org/10.20897/ejsteme/18917>**Published:** July 5, 2026**ABSTRACT**

Artificial Intelligence (AI) is fundamentally altering the landscape of Science, Technology, Engineering, and Mathematics (STEM) education. By shifting from simple procedural automation toward deep semantic reasoning, adaptive personalization, and real-time diagnostic support, AI technologies offer promising opportunities while simultaneously raising critical questions regarding pedagogical validity, cognitive load, data privacy, and systemic equity. This editorial note synthesizes the core contributions of six peer-reviewed papers published in this special issue. Together, these articles explore AI integration across diverse domains, including distance programming environments, PRISMA-based literature reviews, augmented reality in didactics, and inclusive, culturally contextualized STEM instruction. By situating these findings within the broader scholarly discourse indexed in major scientific databases, this note highlights key emergent themes—specifically, practitioner-led co-design, the alignment of pedagogical intent with AI capabilities, and the imperative for ethical, equitable deployment.

Keywords: artificial intelligence, science, technology, engineering, mathematics

The integration of Artificial Intelligence (AI) into Science, Technology, Engineering, and Mathematics (STEM) education represents a paradigm shift in how instructional content is designed, delivered, and assessed (Aldemir et al., 2026). Historically, educational technologies operated primarily as static administrative tools or rigid, rule-based practice platforms. Today, advancements in Large Language Models (LLMs), machine learning algorithms, adaptive tutoring systems, and augmented reality (AR) have enabled dynamic, interactive, and personalized learning environments (Bicer et al., 2026).

However, as technological capabilities accelerate, educational researchers and practitioners face a persistent dilemma: bridging the gap between technological potential and demonstrated classroom effectiveness (Aldemir et al., 2025). The uncritical adoption of AI risks exacerbating existing digital divides, imposing cognitive overburden through ill-designed interfaces, or degrading critical thinking through over-reliance on automated outputs. To

address these challenges, contemporary STEM research must move beyond technical feasibility studies toward rigorous, practitioner-centered, and pedagogically grounded investigations (Bicer et al., 2026).

This special issue of the European Journal of STEM Education titled *AI and STEM: Exploring the Impact of Artificial Intelligence on Science, Technology, Engineering, and Mathematics Education* brings together six distinct contributions that address these themes. The collected papers examine AI and technology-enhanced learning through multiple methodological lenses, including participatory design research (PDR), systematic literature reviews, didactical engineering, and mixed-methods empirical studies. This editorial note outlines the theoretical foundations, empirical findings, and overarching implications of these six studies, grounding their insights within the broader international research landscape.

Overview of contributions to the special issue

The six papers included in this special issue collectively address the multi-faceted challenges and possibilities of modern STEM education.

Customized chatbots to support collective mathematical argumentation

Yuling Zhuang, Wisdom Yao Nudze, Xiangquan Yao, and Trina J. Davis explore the application of customized, GPT-based chatbots to facilitate preservice teachers' understanding of collective mathematical argumentation. By framing the chatbot as a collaborative dialogue partner trained on specific argumentation rubrics (e.g., Toulmin's model), the study demonstrates how tailored LLMs can coach future educators in constructing, critiquing, and facilitating multi-perspective mathematical reasoning and justification.

Static vs. adaptive AI feedback on student self-explanations

Gamze Ozogul, Xiaoying Zheng, Wookhee Min, Seung Lee, Jordan Esiason, Dan Carpenter, and James Lester conduct an empirical investigation comparing student experiences with static versus adaptive AI-generated feedback on self-explanations within introductory computer science courses. By evaluating how personalized, context-aware AI scaffolding impacts learner engagement and problem-solving, their findings demonstrate that adaptive AI feedback significantly improves self-explanation quality, reduces confusion during complex coding tasks, and helps students maintain metacognitive focus compared to generic, static prompts.

Augmented reality for tangible AI/ML concepts

Wolfgang Robinig, Benedikt Br  nner, Harald Burgsteiner, and Johannes Peter Wallner address the challenge of teaching abstract AI and Machine Learning (ML) concepts in K–12 and teacher training. They introduce an interactive Augmented Reality (AR) Mask that projects real-time computer vision feature extraction, classification, and neural network node activations directly onto facial/spatial models. Their work demonstrates how coupling physical tangibility with AR visualizations demystifies complex ML algorithms, lowers cognitive entry barriers, and equips preservice teachers with intuitive didactic strategies for K–12 classrooms.

Semantic code analysis and distance programming education

In their study, Mery Noviyanti, Sudirman Sudirman, Thesa Kandaga, Suci Nurhayati, and Muhamad Galang Isnawan investigate the infrastructural and conceptual barriers facing distance-learning students in computer programming courses. Standard Learning Management System (LMS) plugins, such as Moodle's Virtual Programming Laboratory (VPL), evaluate code solely through binary test-case matching. While effective for verifying functional execution, this mechanism fails to detect underlying conceptual misalignments or evaluate Computational Thinking (CT).

To bridge this gap, they employed a Participatory Design Research (PDR) approach across five phases, engaging course lecturers as co-designers to develop an *AI-Enhanced Python Virtual Laboratory (AI-PVL)* integrating GPT-4 semantic analysis directly into Moodle. Their empirical findings highlight several key insights:

- Needs Analysis (N=65): Identified five critical features with negative gap scores exceeding -2.0 (auto-save, real-time feedback, code visualization, execution, and debugging support), demonstrating that foundational environment stability is a prerequisite for CT development.
- Practitioner-Driven Co-Design: Lecturers co-created three major innovations: a dual-feedback interface (pairing objective test-case results with AI semantic commentary), a cohort-level misconception frequency map for instructors, and co-authored semantic prompt templates tied to specific CT rubrics.
- Validation and Feasibility: Expert review via Aiken's V confirmed high content (90%) and media (90%) validity, yielding an overall feasibility index of 84%.

Mapping the landscape of AI in mathematics education

Providing a macro-level synthesis, Fariza Omirzakova and Sarsenkul Tileubay present a PRISMA-based systematic literature review analyzing 42 peer-reviewed studies published between 2021 and August 2025 on AI applications in K–12 and higher mathematics education. Using the Mixed Methods Appraisal Tool (MMAT), the authors evaluate the shift from traditional intelligent tutoring systems (ITS) to generative AI and LLM-driven personalized feedback. Key insights from their systematic review include:

- **Thematic Shift:** A transition from early technological benchmarking toward investigating teacher readiness, responsible AI design, and academic integrity.
- **Geographic and Disciplinary Trends:** While research output is internationally widespread, disciplinary emphasis remains heavily concentrated in general Education (60%) and Computer Science (28%), with limited representation in humanities-focused or socio-ethical dimensions.
- **Evidence Heterogeneity:** The authors warn against equating technical sophistication with classroom efficacy, highlighting that long-term, adequately powered empirical studies remain scarce.

Where practical benefits meet ethical vigilance: Educators and AI

Examining practitioner-level dynamics, Ela Suryani, Kartika Yuni Purwanti, Mochamad Rizqi Adhi Pratama, and Wulan Aulia Azizah conducted a quantitative survey evaluating how certified educators (N = 218) in Semarang Regency, Central Java, Indonesia, perceive the intersection of artificial intelligence's utility and its associated ethical risks. Utilizing a purposive sampling strategy, data gathered via a structured questionnaire were evaluated through descriptive metrics and Pearson's product–moment correlation. The authors highlighted three primary findings:

- **Exposure vs. Conceptual Understanding:** A majority of respondents (64%) reported hands-on experience with AI tools, whereas a considerably smaller proportion (27%) expressed confidence in their comprehension of foundational AI principles. The authors note that while these baseline figures suggest a broader trend where practical exposure exceeds conceptual familiarity across the sample, they do not empirically demonstrate an individual-level deficit.
- **Positive Perceptions alongside Ethical Vigilance:** The educators generally affirmed AI's operational value (M = 4.18, SD = 0.63) while simultaneously maintaining a strong critical awareness of ethical concerns (M = 3.42, SD = 0.57), specifically regarding algorithmic accuracy, data privacy, systemic bias, transparency, institutional accountability, and the necessity of human oversight.
- **Complementary Constructs:** A moderate, statistically significant positive correlation was identified between operational appreciation and ethical vigilance, $r(216) = 0.49$, 95% CI [0.38, 0.59], $p < .001$. This relationship indicates that educators who recognize AI's pedagogical utility are also more attuned to its underlying risks, emphasizing the need for professional development initiatives that integrate practical training with critical AI literacy.

Synthesizing the contributions of these six studies alongside wider literature indexed in Scopus reveals three overarching themes that define the current state of AI and STEM education research.

Practitioner-centered co-design and co-invention

A prominent theme across the special issue is the necessity of involving educators directly as co-designers rather than passive end-users of technology. Mery Noviyanti, Sudirman Sudirman, Thesa Kandaga, Suci Nurhayati, and Muhamad Galang Isnawan demonstrate that involving lecturers in system design yielded critical features, such as the misconception map and dual-feedback display, that technical developers alone could not have anticipated.

This aligns with findings in design-based implementation research (DBIR), which emphasize that sustainable technology integration depends on aligning tools with existing institutional workflows and practitioners' pedagogical content knowledge (Fishman et al., 2013; Awidi & Paynter, 2024). When AI systems are engineered without teacher input, they frequently suffer from adoption resistance or pedagogical misalignment (Bicer et al., 2026).

To ensure meaningful co-design, educators must first possess a foundational understanding of underlying AI mechanisms (Aldemir et al., 2025). Wolfgang Robinig, Benedikt Br  nner, Harald Burgsteiner, and Johannes Peter Wallner address this prerequisite by evaluating an augmented reality (AR) mask activity designed to navigate learners through an end-to-end machine learning workflow within a single instructional session. In this activity, participants collect and annotate facial expression data, train an image classification model via transfer learning, and integrate it into a Scratch-based AR application driven by real-time webcam input.

An exploratory mixed-methods study involving 32 in-service teachers and 40 lower-secondary students yielded the following key findings:

- **Instructional Efficacy and Pedagogical Alignment:** In-service teachers evaluated the module as an effective vehicle for introducing core computer science concepts, including neural network training, computer vision principles, algorithmic error mechanics, and dataset bias.
- **Student Engagement and Conceptual Transfer:** Lower-secondary students reported high engagement and successfully articulated the direct relationship between model accuracy and the size and diversity of the underlying training dataset.
- **Transition to Informed Co-Designers:** By providing low-threshold, experiential engagement with AI fundamentals, such scaffolding tools facilitate a key professional shift, empowering educators to move from passive consumers of educational technology to critical, informed partners in AI design and implementation.

Aligning AI feedback with cognitive load theory

The way students and teachers process information is directly shaped by effective designs of AI feedback mechanisms (Bicer et al., 2026). This has been illustrated by Yuling Zhuang, Wisdom Yao Nudze, Xiangquan Yao, and Trina J. Davis in the way preservice teachers understand collective mathematical argumentation through customized ChatGPT-based chatbots. As noted by Mery Noviyanti, Sudirman Sudirman, Thesa Kandaga, Suci Nurhayati, and Muhamad Galang Isnawan, environment setup failures and cryptic compiler errors impose significant extraneous cognitive load, suppressing the germane load required for algorithmic schema construction. This insight builds directly upon Sweller et al.'s (2019) Cognitive Load Theory.

Investigating the scalable delivery of AI-driven pedagogical support, Gamze Ozogul, Xiaoying Zheng, Wookhee Min, Seung Lee, Jordan Esiason, Dan Carpenter, and James Lester conducted a comparative case study across two consecutive semesters of an undergraduate computer science course to evaluate classroom response systems providing either static, instructor-authored feedback or adaptive feedback generated via natural language processing on student self-explanations. While both modalities effectively maintained student engagement and course participation, the adaptive platform demonstrated the capacity of automated systems to scale formative assessment in large-enrollment contexts, foster deeper conceptual mastery, and reduce instructional burden. Nevertheless, the authors emphasize that enhancing algorithmic accuracy remains critical to mitigate the risk of students developing an uncritical reliance on automated feedback.

Recent empirical research in AI-assisted programming likewise confirms that while LLM-generated hints can accelerate error resolution, feedback must be carefully scaffolded to prevent over-reliance and promote metacognitive reflection (Pankiewicz & Baker, 2024; Aldemir et al., 2025a; Kinder et al., 2025). The dual-feedback mechanism highlighted in this issue ensures that students maintain a grounded "credibility anchor" through objective code execution while engaging with higher-level semantic diagnostics.

Ethical guardrails, data governance, and digital equity

As AI tools become ubiquitous, questions of data privacy, algorithmic bias, and equitable access become paramount. Fariza Omirzakova and Sarsenkul Tileubay emphasize that digital divides persist both globally and locally, threatening to disadvantage student populations with limited digital infrastructure or restricted access to commercial AI APIs.

Empirical evidence indicates that educators actively engage with the ethical implications of artificial intelligence rather than remaining passive observers. In a survey of 218 certified educators, Ela Suryani, Kartika Yuni Purwanti, Mochamad Rizqi Adhi Pratama, and Wulan Aulia Azizah observed a positive, moderate correlation ($r = 0.49$) between perceived operational utility and heightened critical awareness regarding dataset accuracy, privacy, algorithmic bias, transparency, institutional accountability, and human oversight. Rather than acting as opposing forces, practical enthusiasm and ethical vigilance function as complementary facets of professional adoption; however, because widespread hands-on tool usage (64%) markedly outpaced confidence in foundational AI mechanics (27%), a distinct need emerges for professional development that pairs conceptual literacy with classroom implementation—an objective directly addressed by experiential learning tools such as the augmented reality mask module designed by Wolfgang Robinig, Benedikt Brünner, Harald Burgsteiner, and Johannes Peter Wallner.

To maintain institutional integrity, AI deployments must implement strict data minimization, purpose limitation, and anonymization protocols. As highlighted by Raza et al. (2025) and Zacharis and Papadakis (2025), safeguarding academic integrity and ensuring construct validity in automated assessment are essential prerequisites for ethical educational technology design. Future directions must also explore open-weight, locally hosted LLMs to eliminate third-party data transmission and bridge infrastructure gaps for resource-constrained institutions (Koutchme et al., 2024).

Conclusion and Future Research Agenda

The research collected in this special issue demonstrates that Artificial Intelligence and interactive technologies hold transformative potential for STEM education (Bicer et al., 2026). However, technological capability must be guided by sound pedagogical principles, robust empirical validation, and ethical commitments to equity. To advance this field, we outline three priority directions for future scholarship:

- Longitudinal and Adequately Powered Experimental Studies: Transitioning from proof-of-concept pilots to randomized, large-sample studies that track long-term conceptual retention and skill transfer across semesters.
- Investigation of Open-Source / Local LLM Architectures: Comparative evaluations of reliance on commercial API dependency against lightweight, open-weight local LLM deployments to enhance data sovereignty and lower costs for developing institutions.
- Analytics on Metacognitive Engagement: Utilizing fine-grained learning analytics—such as revision histories and time-on-task metrics—to examine how AI feedback shapes student problem-solving behavior over time.

We extend our sincere gratitude to all the authors, reviewers, and editorial staff whose dedication made this special issue possible. It is our hope that these contributions will inspire continued interdisciplinary collaboration among researchers, educators, and technology developers in building effective, equitable STEM learning environments.

Biographical sketch



Ali Bicer is an Associate Professor of Mathematics Education in the Department of Teaching, Learning and Culture at Texas A&M University, United States, where he earned his Ph.D. in Curriculum and Instruction. Before returning to Texas A&M in 2024, he earned tenure at the University of Wyoming. His research focuses on mathematical creativity, problem posing and solving, STEM education, teacher professional development, and artificial intelligence in mathematics teaching and learning. He is principal investigator of a five-year, \$2 million NSF IUSE project on AI-guided professional learning for preservice mathematics teachers and directs the Creative Mathematics Teaching and Learning Lab (CMTL2).



Stamatios Papadakis is an Associate Professor at the Department of Preschool Education, University of Crete, Greece. He holds a PhD in Educational Technology. His research focuses on artificial intelligence in education, educational technology, mobile learning, STEM education, and the integration of digital technologies into early childhood and teacher education.



Dr. Tugce Aldemir is an Assistant Professor of STEM Education in the Department of Teaching, Learning and Culture at Texas A&M University and principal investigator of the CoHALE Lab. Her research integrates the learning sciences, computer science, and instructional design to create technology-supported learning experiences that broaden participation in STEM and are sustainable in schools. Through design-based research and mixed methods, she collaborates with educators to develop learning environments, examine learning in authentic settings, and generate design principles and theoretical insights. She earned her Ph.D. in Learning, Design, and Technology from The Pennsylvania State University and completed postdoctoral work at the University of Pennsylvania and the University of Connecticut.



Dr. Ariffin Abdul Mutalib is a UX Professor at Universiti Utara Malaysia, Malaysia. He obtained his master's degree in interactive Multimedia (HCI) in 2001 from Heriot-Watt University, Scotland. Since then, he has been applying Design Thinking in his research work. He has successfully supervised more than 10 PhD projects related to UX, Usability, and multimedia in education, most of which have received academic and industry awards. He has been awarded a few titles internationally. Now, he is actively mentoring academics as well as communities in various aspects, by applying UX techniques in his delivery.

REFERENCES

- Aldemir, T., Bicer, A., Kilinc, S., Moon, J., & Kwok, M. (2025a). Exploring emergent AI-TPACK competencies in a two-week AI literacy module for preservice teachers. *Teaching and Teacher Education*, 168, 105231. <https://doi.org/10.1016/j.tate.2025.105231>
- Aldemir, T., Kilinc, S., Bicer, A., Grant, P., Davis, T., & Sweany, N. W. (2025). Intelligent-TPACK in practice: design and evidence from a three-week teacher preparation module. *Computers and Education Open*, 100306. <https://doi.org/10.1016/j.caeo.2025.100306>
- Aldemir, T., Bicer, A., Kilinc, S., Moon, J., & Kwok, M. (2026). Challenges, safeguards, and professional learning needs for AI integration: insights from a two-week AI literacy module with preservice teachers. *Cogent Education*, 13(1), 2721038. <https://doi.org/10.1080/2331186X.2026.2721038>

- Awidi, I. T., & Paynter, M. (2024). An evaluation of the impact of digital technology innovations on students' learning: Participatory research using a student-centred approach. *Technology, Knowledge and Learning*, 29, 65–89. <https://doi.org/10.1007/s10758-022-09619-5>
- Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National Society for the Study of Education Yearbook*, 112(2), 136–156.
- Kinder, A., Briese, F. J., Jacobs, M., Dern, N., Glodny, N., Jacobs, S., & Leßmann, S. (2025). Effects of adaptive feedback generated by a large language model: A case study in teacher education. *Computers and Education: Artificial Intelligence*, 8, Article 100349. <https://doi.org/10.1016/j.caeai.2024.100349>
- Koutcheme, C., Dainese, N., Sarsa, S., Hellas, A., Leinonen, J., & Denny, P. (2024). Open-source language models can provide feedback: Evaluating LLMs' ability to help students using GPT-4-as-a-judge. In *Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education* (Vol. 1, pp. 52–58). Association for Computing Machinery. <https://doi.org/10.1145/3649217.3653612>
- Pankiewicz, M., & Baker, R. S. (2024). Navigating compiler errors with AI assistance: A study of GPT hints in an introductory programming course. In *Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education* (Vol. 1, pp. 94–100). Association for Computing Machinery. <https://doi.org/10.1145/3649217.3653608>
- Raza, F. A., Singh, A. D., Kovilpillai, J. J. S., Hamdan, A., & Rajaratnam, V. (2025). Safeguarding integrity in AI-enhanced education: Stakeholder perspectives on accuracy, validity, and ethics in ASEAN. *European Journal of STEM Education*, 10(1), Article 22. <https://doi.org/10.20897/ejsteme/17307>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Zacharis, G., & Papadakis, S. (2025). Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment. *Educational Process: International Journal*, 19, Article e2025591. <https://doi.org/10.22521/edupij.2025.19.591>