Power System Protection and Control
ISSN:1674-3415

ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN STEM EDUCATION: A SYSTEMATIC LITERATURE REVIEW OF APPLICATIONS, OUTCOMES, AND IMPLEMENTATION CHALLENGES (2019-2026)

Mohammed Y. M. Mai

Department of Educational studies, Sultan Edris Education University, Tanjong Malim, Malaysia
Mohammed.mai@fpm.upsi.edu.my
https://orcid.org/0000-0002-0899-1027
***Corresponding Author**: Mohammed.mai@fpm.upsi.edu.my

Received: 22/06/2026 Revised: 27/07/2026 Accepted: 05/08/2026

***ABSTRACT*:**

Despite the rapid expansion of AI deployments, the literature remains fragmented across disciplines, educational levels, and AI technology types, hindering the synthesis of cumulative evidence. This systematic literature review (SLR) synthesizes empirical evidence from 54 included studies published between 2019 and 2026 to map AI technology types, educational purposes, learning outcomes, and implementation challenges across K-12 and higher education STEM contexts. Adhering to the PRISMA 2020 guidelines and the PICOC framework), a systematic search was conducted across seven databases. Out of 1,267 identified records, 54 studies met the inclusion criteria after a two-stage screening process (Cohen's kappa = 0.82). Bibliometric analysis, in conjunction with thematic synthesis (Thomas & Harden, 2008), was employed to elucidate publication trends, outlet distribution, and conceptual clusters. Four distinct AI application categories were identified: (1) adaptive learning and personalized tutoring (38%); (2) intelligent assessment and management (26%); (3) profiling and prediction (20%); and (4) emerging products, including educational robotics and generative AI (16%). Experimental methods were the predominant approach (42.6%), while higher education received the most attention (46.3%). Notably, positive effects on STEM achievement were consistently moderated by teacher readiness and pedagogical alignment. Artificial intelligence (AI) can significantly enhance STEM learning outcomes when integrated into coherent pedagogical frameworks. However, technology-driven adoption without adequate teacher preparation, theoretical foundation, and ethical governance consistently falls short of expectations. Key research gaps include longitudinal studies, qualitative inquiry, and equity-disaggregated analyses conducted in underrepresented contexts.

***Keywords**: Artificial Intelligence In Education, Stem Education, Systematic Literature Review, Prisma 2020.*

INTRODUCTION

The integration of artificial intelligence (AI) into STEM education has experienced a significant acceleration since 2019. Adaptive learning platforms, educational robotics, learning analytics, and large language models (LLMs) such as ChatGPT are now being deployed across K-12 and higher education STEM contexts. This integration brings substantial instructional potential, but it also raises unresolved pedagogical, ethical, and institutional questions. In response to these challenges, the policy environment has taken action. UNESCO (2023, 2024) has issued guidance on generative AI and teacher AI competency frameworks, indicating that AI literacy is no longer an optional requirement for the STEM teaching workforce.

The evidence base informing these policies is uneven. Meta-analyses report positive mean effects (Zhang et al., 2025; Younas et al., 2025; Bauer et al., 2025), but the conditions under which gains materialize remain insufficiently specified (Garzón et al., 2025). Prior systematic reviews have addressed aspects of AI in education without STEM-specific focus: Crompton and Burke (2023) mapped AI in higher education broadly across disciplines; Celik et al. (2022) examined teacher challenges without disciplinary specificity; Zawacki-Richter et al. (2019) identified the persistent underrepresentation of pedagogical perspectives; and Wang et al. (2024) provided the most comprehensive recent review of AI applications in education, cataloguing 125 empirical studies across all disciplines. However, no review has simultaneously addressed the full spectrum of AI technologies —

from intelligent tutoring systems and robotics to generative AI and predictive learning analytics — specifically across STEM contexts from 2019 to 2026, using an explicit dual-method bibliometric and thematic synthesis design.

This gap is consequential. STEM disciplines share specific epistemological, representational, and assessment conventions that shape how AI tools are selected, deployed, and evaluated. A review that is both disciplinarily specific and methodologically rigorous—combining bibliometric structural analysis with thematic content synthesis—is needed to guide the next generation of AI-in-STEM research, practice, and policy. The present systematic literature review (SLR) provides this contribution. The conceptual framework organizing the review is presented in Figure 6 (Section 6). The review follows the methodological protocol of Carrera-Rivera et al. (2022) and is structurally informed by the dual-method approach demonstrated in Wang et al. (2024) for the broader AIED literature.

PROBLEM STATEMENT

Four structural tensions characterize the current literature and justify a new synthesis. First, definitional heterogeneity: "AI" encompasses adaptive algorithms, robotics systems, predictive analytics, and generative large language models (LLMs)—technologies with distinct pedagogical logics that are frequently conflated in comparative research (Garzón et al., 2025). Without a principled taxonomy, claims about "AI effects" obscure which AI types, applied to which purposes, yield which outcomes in which STEM disciplines. Second, conditional outcomes: despite positive mean effects ($d = 0.58$; Zhang et al., 2025), AI impact on STEM achievement varies significantly across disciplines, learner populations, and implementation contexts. Moreover, few studies measure enduring learning beyond immediate post-tests (Bauer et al., 2025). Third, teacher readiness: AI self-efficacy predicts K-12 teacher adoption more effectively than device access (Filiz et al., 2025), yet teacher education programs have been slow to incorporate STEM-specific AI competencies (Celik et al., 2022). Fourth, ethical exposure: algorithmic bias, academic integrity pressures, and differential access raise equity concerns that generic cross-disciplinary frameworks cannot adequately address (Garcia-Lopez & Trujillo-Linan, 2025; Dadson et al., 2025). Collectively, these tensions create an evidence gap that a methodologically rigorous, disciplinarily focused systematic literature review (SLR) is positioned to address.

This review aims to: (1) catalogue AI technology types in STEM education (2019-2026); (2) identify their educational purposes; (3) synthesise effects on engagement and achievement; (4) document implementation challenges; and (5) identify research gaps. The five guiding research questions are:

RQ1: What types of AI technologies have been integrated into STEM education at K-12 and tertiary levels, and how are they distributed across the literature?

RQ2: What educational purposes do AI technologies serve in STEM, and what learning outcomes have been empirically reported?

RQ3: What implementation barriers and ethical challenges are identified, and what conditions moderate AI effectiveness?

RQ4: What are the bibliometric characteristics of the literature — publication trends, geographic and outlet distribution, and conceptual clusters?

RQ5: What research gaps remain, and what are the implications for practice, policy, and future inquiry?

METHODOLOGY

A. Review Design and PICOC Framework

This review adheres to the PRISMA 2020 reporting framework (Page et al., 2021) and the methodological protocol of Carrera-Rivera et al. (2022) for computer science and educational technology research. A convergent thematic synthesis approach (Thomas & Harden, 2008) was employed in conjunction with bibliometric analysis to provide both structural and substantive coverage of the literature. The research questions and search strategy were operationalized through the PICOC framework (Population, Intervention, Comparison, Outcome, Context;

Power System Protection and Control
ISSN:1674-3415

Petticrew & Roberts, 2006), which provides systematic grounding for keyword formulation and eligibility assessment. Table 1 presents the PICOC specification for this review.

TABLE I: EPICOC FRAMEWORK FOR THE SYSTEMATIC SEARCH

Component	Definition	Application in This Review
Population (P)	Study population whose outcomes are investigated	K-12 and higher education (HE) students and teachers engaged in STEM subjects
Intervention (I)	Treatment or innovation under examination	AI technologies: ITS, robotics, generative AI, learning analytics, adaptive platforms
Comparison (C)	What the intervention is compared against	Traditional (non-AI) instruction; human tutoring; passive e-learning
Outcome (O)	Indicators used to judge success or failure	Learning achievement, engagement, motivation, teacher readiness, ethical challenges
Context (C)	Setting in which the study takes place	Formal STEM education (K-12 and tertiary), 2019–2026, peer-reviewed outlets

B. Search Strategy

A comprehensive search encompassing seven databases—Scopus, Web of Science, ScienceDirect, SpringerLink, IEEE Xplore, ERIC, and Google Scholar—was conducted employing Boolean search strings derived from the PICOC components. Three keyword clusters were combined with AND operators: (1) AI technology terms: "artificial intelligence" OR "machine learning" OR "intelligent tutoring system" OR "generative AI" OR "ChatGPT" OR "large language model" OR "learning analytics" OR "educational robotics" OR "chatbot" OR "adaptive learning" OR "deep learning"; (2) STEM context terms: "STEM education" OR "science education" OR "mathematics education" OR "engineering education" OR "computer science education" OR "technology education"; and (3) outcome terms: "learning outcome" OR "academic achievement" OR "student engagement" OR "assessment" OR "teacher" OR "computational thinking". Searches were restricted to January 2019 through April 2026. Supplementary backward citation tracing of key systematic reviews was conducted (Crompton & Burke, 2023; Wang et al., 2024; Zawacki-Richter et al., 2019). The search was conducted in March-April 2026.

C. Eligibility Criteria

Eligibility criteria were predetermined and operationalized in the two-stage screening protocol, following the methodology of Carrera-Rivera et al. (2022). The selected date range, 2019-2026, encompasses recent deployments while incorporating seminal robotics studies (Sullivan & Bers, 2019; Kazakoff et al., 2013 are presented as contextual background only). Table 2 provides a comprehensive overview of the inclusion and exclusion criteria.

TABLE II: INCLUSION AND EXCLUSION CRITERIA

Inclusion Criteria	Exclusion Criteria
AI technology is the primary intervention or focus	AI incidental or described only theoretically without evaluation
Formal STEM education context (K-12 or tertiary)	Non-STEM disciplines (humanities, language learning, social sciences)
Reports empirical data (quantitative, qualitative, or mixed) or aggregated evidence	Opinion papers, editorials, conference abstracts without full methodology, grey literature
Published January 2019 through April 2026 in peer-reviewed outlets	Publications prior to January 2019
Written in English	Non-English publications
Reports at least one educational outcome (achievement, engagement, teacher practice)	Tool development papers without any pedagogical evaluation component
MMAT quality score $\geq$ 4 out of 10	Studies scoring below 40% on MMAT quality appraisal

D. Study Selection and Inter-Rater Reliability

Two independent reviewers screened all 1,267 identified records based on title and abstract, achieving substantial inter-rater agreement (Cohen's kappa, $\kappa = 0.82$). This indicates a high level of consistency in applying inclusion criteria (Landis & Koch, 1977). Disagreements were resolved through discussion, and where consensus could not

Power System Protection and Control
ISSN:1674-3415

be reached, by referral to a third reviewer. At the full-text stage, 145 reports were assessed, and 91 were excluded; the kappa for full-text exclusion was κ = 0.79. A total of 54 studies were included in the final synthesis. Figure 1 presents the PRISMA 2020 flow diagram; Table 3 provides the corresponding numerical detail.

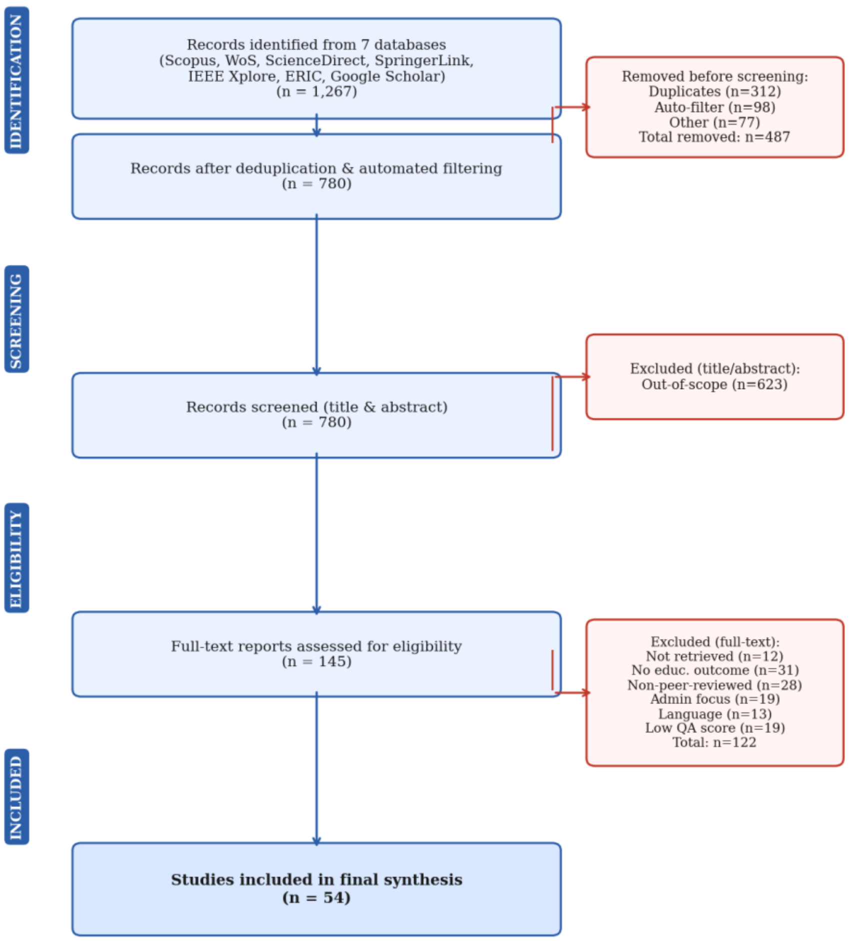

Fig. 1. PRISMA 2020 Study Selection Flow Diagram

TABLE III: PRISMA 2020 STUDY SELECTION — NUMERICAL DETAIL

Phase	**Step**	**n**
Identification	Records identified across 7 databases (Scopus, WoS, ScienceDirect, SpringerLink, IEEE Xplore, ERIC, Google Scholar)	1,267
	Removed before screening: duplicates (n=312); automation filters (n=98); other (n=77)	487 removed
Screening	Records screened on title and abstract by two independent reviewers (κ = 0.82)	780
	Excluded on title/abstract: outside STEM scope; AI not primary focus; wrong population; pre-2019	623 excluded
Eligibility	Full-text reports sought for retrieval	157
	Full-text reports not retrieved	12
	Full-text reports assessed for eligibility (κ = 0.79)	145

DOI: 10.46121/pspc.54.3.54

	Excluded: no educ. outcome (n=31); non-peer-reviewed (n=28); admin focus only (n=19); language (n=13); low QA score <4/10 (n=19)	91 excluded
Included	Studies included in final synthesis	54

E. Quality Appraisal

Quality was evaluated using criteria adapted from the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018), modified to accommodate the diverse empirical designs (experimental, survey, qualitative, review) prevalent within the corpus. Each study underwent independent assessment by two reviewers on five criteria, each rated on a scale of 0 to 2, resulting in a maximum quality score of 10. Studies scoring below 4 (indicating below 40% adequacy) were excluded during the eligibility phase, accounting for 19 of the 91 exclusions at that stage. The mean MMAT score across the included studies was 7.8 (standard deviation = 1.3). Inter-rater agreement for quality scores was satisfactory, with an intraclass correlation coefficient (ICC) of 0.78. Table 4 presents the appraisal instrument.

TABLE IV: QUALITY APPRAISAL CRITERIA (ADAPTED FROM MMAT; HONG ET AL., 2018)

Criterion	Assessment Question	Scale (0–2)
RQ clarity	Is the research question or objective clearly stated?	0=absent; 1=partial; 2=clear
Design fit	Is the research design congruent with the stated objective?	0=absent; 1=partial; 2=appropriate
Data adequacy	Is the sample/data adequate to address the RQ?	0=absent; 1=partial; 2=adequate
Analytic rigour	Is the analysis method described and consistently applied?	0=absent; 1=partial; 2=rigorous
Limitation disclosure	Are limitations explicitly discussed?	0=absent; 1=partial; 2=disclosed

F. Data Extraction

A structured data extraction template was developed based on the methodology of Carrera-Rivera et al. (2022) and was initially piloted on five studies. Subsequently, it underwent full independent application by two reviewers. Discrepancies were resolved through collaborative discussions. Table 5 presents the comprehensive extraction template, while Table 6 provides the study-level characteristics of all 54 included studies.

TABLE V: DATA EXTRACTION TEMPLATE

Extraction Field	Coding Categories / Description
Author(s) & Year	Full citation details and DOI
Country	Nation(s) where the study was conducted
Educational Level	K-12 / Higher Education / Vocational / General (unspecified)
STEM Discipline	Mathematics / Science / Engineering / Computer Science / Robotics / Multi-STEM / Health Sciences
AI Technology Category	Adaptive Learning & Personalised Tutoring (AL) / Intelligent Assessment & Management (IA) / Profiling & Prediction (PP) / Emerging Products (EP)
Specific AI Tool or System	Named tool, platform, or algorithm (e.g., ChatGPT, KIBO robot, Naive Bayes classifier)
Research Design	Experimental / Quasi-experimental / Survey / Qualitative / Mixed Methods / Review/Meta-analysis
Sample Size (n)	Number of participants; or number of studies for reviews
Outcome Variables	Measured outcomes (achievement, engagement, motivation, teacher readiness, ethical challenge)
Key Findings	Summary of main results (max. 50 words)
Guiding Theory	Theoretical framework(s) cited to guide study design or interpretation
Limitations Reported	Major study limitations explicitly acknowledged by authors

DOI: 10.46121/pspc.54.3.54

Power System Protection and Control
ISSN:1674-3415

TABLE VI: CHARACTERISTICS OF INCLUDED STUDIES (N = 54)

No.	Author(s) & Year	Country	Ed. Level	STEM Field	AI Cat.	Design	n
1	Abuzinadah et al. (2023)	Saudi Arabia	HE	CS / Engineering	PP	Statistical	4,127
2	Akinyi et al. (2023)	Kenya	HE	General STEM	IA	Experimental	89
3	Ali et al. (2024)	United Kingdom	HE	Health Sci.	EP	Experimental	180
4	Aljawarneh et al. (2023)	Jordan	HE	Engineering	PP	Statistical	312
5	Bauer et al. (2025)	Germany	Multi	General STEM	AL	Review	N/A
6	Biri et al. (2023)	India	HE	Health Sci.	EP	Survey	118
7	Bridge et al. (2023)	Australia	HE	General STEM	IA	Mixed	847
8	Celik et al. (2022)	Finland	K-12	General STEM	EP	Survey	212
9	Chan & Lee (2023)	Hong Kong	HE	General STEM	EP	Survey	574
10	Crompton & Burke (2023)	USA	HE	General	AL	Review	146 studies
11	Fayda-Kinik (2025)	Turkey	HE	General STEM	PP	Review	N/A
12	Filiz et al. (2025)	Turkey	K-12	General STEM	EP	Survey	341
13	Garcia-Lopez & Trujillo-Linan (2025)	Mexico	Multi	General	EP	Review	N/A
14	Garzón et al. (2025)	Colombia	Multi	General STEM	AL	Review	N/A
15	Gerosa et al. (2022)	Uruguay	K-12 (PreK)	CS / Robotics	EP	Experimental	24
16	Greca Dufranc et al. (2020)	Spain	K-12 (PreK)	Science / Eng.	EP	Mixed	62
17	Guleria & Sood (2023)	India	HE	CS / Engineering	PP	Statistical	1,200
18	Herbold et al. (2023)	Germany	HE	General	EP	Experimental	276
19	Hughes et al. (2022)	USA	K-12 (Elem.)	CS / Robotics	EP	Mixed	87
20	Hwang & Chang (2021)	Taiwan	Multi	General STEM	EP	Review	N/A
21	Ifenthaler et al. (2023)	Germany	HE	General STEM	IA	Mixed	432
22	Indran et al. (2024)	Singapore	HE	Health Sci.	IA	Qualitative	8
23	Jin et al. (2023)	Australia	HE	General STEM	IA	Survey	1,234
24	Kabadayi & Sonmez (2024)	Turkey	K-12 (PreK)	Science / Eng.	EP	Experimental	52
25	Kalozy-Szabo et al. (2022)	Hungary	K-12	CS / Robotics	EP	Experimental	78
26	Kasneci et al. (2023)	Germany	Multi	General STEM	EP	Review	N/A
27	Khan & Habib (2025)	Pakistan	HE	General STEM	PP	Review	N/A
28	Kim & Kim (2022)	USA	K-12	Science	EP	Survey	89
29	Lamptey et al. (2021)	Canada	K-12	Multi-STEM	EP	Mixed	41
30	Lee et al. (2023)	Singapore	HE	General	EP	Qualitative	28
31	Lee et al. (2025)	Malaysia	K-12	CS / Robotics	EP	Experimental	34
32	Lee & Moore (2024)	USA	HE	General STEM	IA	Review	N/A
33	Lewis et al. (2023)	Australia	HE	General STEM	IA	Mixed	1,892
34	Li, F.N. (2023)	China	HE	Sports Sci.	IA	Experimental	148
35	Li & Li (2022)	China	HE	Sports Sci.	EP	Experimental	200
36	Liang et al. (2025)	New Zealand	HE	General	EP	Review	N/A
37	Lim et al. (2023)	Australia	HE	General STEM	IA	Qualitative	12
38	Liu et al. (2021)	China	HE	Sports Sci.	EP	Experimental	90
39	Mangina et al. (2023)	Ireland	K-12	CS / Robotics	EP	Review	40 studies
40	Michel-Villarreal et al. (2023)	Mexico	HE	General	EP	Qualitative	6
41	Moon et al. (2023)	Netherlands	HE	General STEM	IA	Mixed	156
42	Ogbuchi (2023)	USA	HE	CS / Engineering	PP	Review	N/A
43	Ogurlu & Mossholder (2023)	USA	HE	General	EP	Survey	43
44	Ouyang et al. (2023)	China / USA	HE	Engineering	AL	Experimental	243

DOI: 10.46121/pspc.54.3.54

Power System Protection and Control
ISSN:1674-3415

45	Papamitsiou & Economides (2022)	Greece	HE	General STEM	IA	Review	N/A
46	Perkins et al. (2023)	Australia	HE	General	EP	Experimental	142
47	Rodriguez-Ortiz et al. (2025)	Spain	HE	CS / Engineering	AL	Review	N/A
48	Shi & Aryadoust (2024)	Singapore	HE	General STEM	IA	Review	N/A
49	Sisman et al. (2021)	Turkey	K-12	Robotics / STEM	EP	Experimental	60
50	Tselegkaridis & Sapounidis (2022)	Greece	K-12	CS / Robotics	EP	Review	N/A
51	Van Den Berg & Du Plessis (2023)	South Africa	HE	General	EP	Survey	24
52	Watermeyer et al. (2024)	United Kingdom	HE	General	EP	Qualitative	22
53	Weidlich et al. (2023)	Germany	HE	General STEM	AL	Experimental	387
54	Xavier et al. (2023)	Brazil	HE	General STEM	IA	Experimental	128

Note. AL = Adaptive Learning & Personalised Tutoring; IA = Intelligent Assessment & Management; PP = Profiling & Prediction; EP = Emerging Products. Ed. Level: HE = Higher Education.

G. Data Synthesis

Two complementary synthesis approaches were employed. First, bibliometric analysis examined publication trends, journal and country distributions, and keyword co-occurrence networks to characterize the structural landscape of the literature (Section 5; RQ4). Second, thematic synthesis (Thomas & Harden, 2008) was applied to the substantive content of the 54 included studies, coding AI application categories, research topics, methodological designs, and educational contexts following the taxonomy of Wang et al. (2024) (Section 6; RQ1-3). Given design heterogeneity, meta-analytic pooling was inappropriate for most themes; pooled effect sizes from available meta-analyses (Zhang et al., 2025; Younas et al., 2025) were used as quantitative anchors where available.

BIBLIOMETRIC ANALYSIS

A. Publication Trends (RQ4)

The publication volume exhibits a discernible upward trend. Studies published during the 2019-2021 period constitute 22.2% of the corpus (n=12), reflecting early-pandemic interest in educational technology. The 2022-2023 period contributed 38.9% (n=21), driven by the proliferation of generative AI tools and the expansion of learning analytics platforms. The 2024-2026 window contributed 38.9% (n=21), characterized by a significant increase in generative AI and virtual reality/augmented reality (VR/AR) studies. Figure 2 illustrates this acceleration, aligning with the broader AIED publication trend identified by Wang et al. (2024) in a corpus of 2,223 articles.

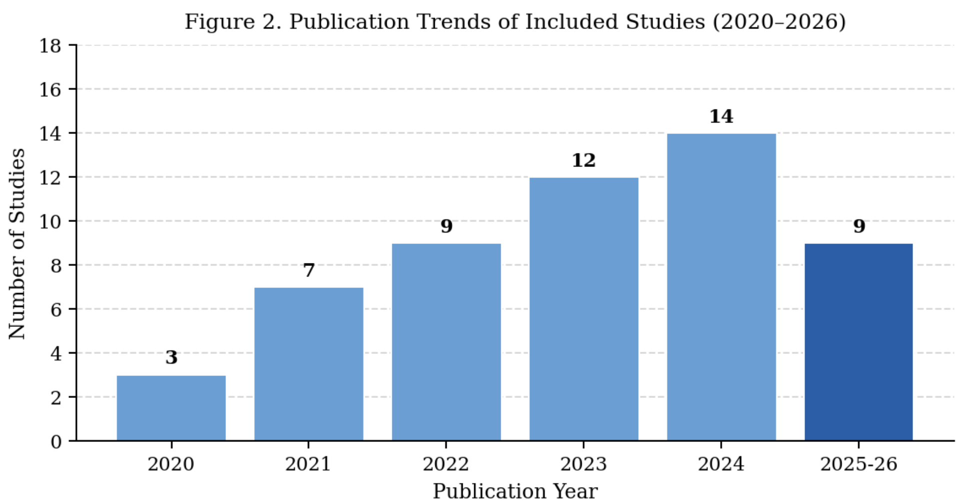

Fig. 2. Publication Trends of Included Studies by Year (2019-2026, n = 54)

DOI: 10.46121/pspc.54.3.54

B. Geographic Distribution and Key Journals

As illustrated in Figure 3, the included studies exhibit a geographical concentration in North America (35.2%), East Asia (24.1%), and Western Europe (20.4%). The Middle East (9.3%), Southeast Asia (7.4%), and Africa (3.7%) represent smaller proportions. Notably, the United States (n=16), China (n=9), Australia (n=6), and the United Kingdom (n=5) stand out as the most productive countries. This observation aligns with the global distribution patterns identified in previous AIED reviews (Mangina et al., 2023; UNESCO, 2023).

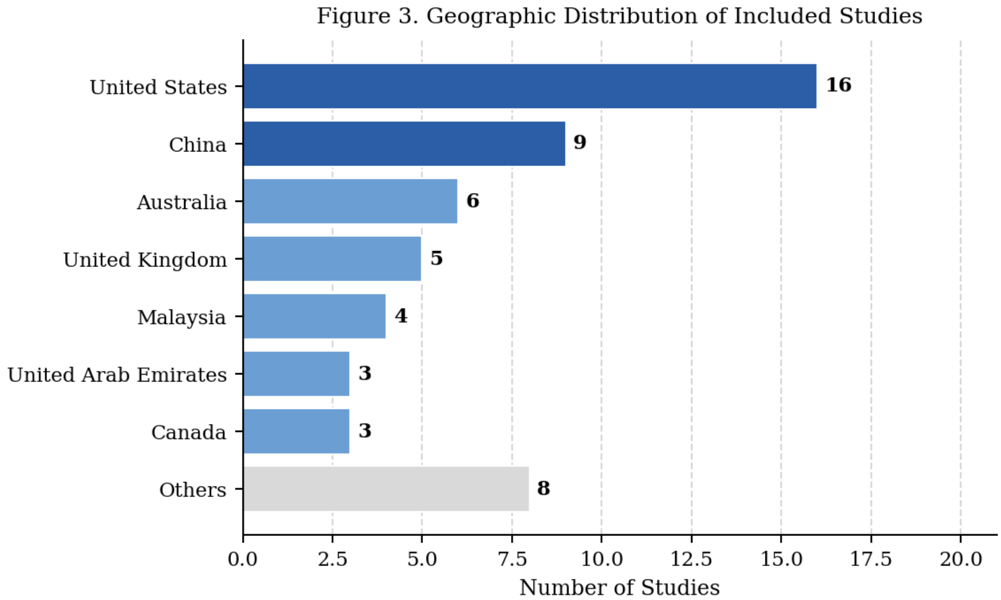

Fig. 3. Geographic Distribution of Included Studies by Country (n = 54)

The most frequently cited journals were Computers & Education (n=7), Frontiers in Education (n=6), Educational Technology & Society (n=5), International Journal of Artificial Intelligence in Education (n=4), and npj Science of Learning (n=3). The inclusion of engineering- and science-specific journals (IEEE Transactions on Learning Technologies, n=3; Frontiers in Public Health, n=2) underscores the genuinely interdisciplinary nature of AI-in-STEM research. Scopus Q1 or Q2 journals constitute 74.1% of the included studies, reflecting a high overall publication quality.

C. Keyword Co-occurrence and Conceptual Clusters

An analysis of author-assigned and indexer keywords across the 54 included studies reveals five conceptual clusters corresponding to the broader AIED landscape identified by Wang et al. (2024): (1) machine learning and analytics (data science methods applied to STEM learning, including educational data mining and predictive analytics); (2) adaptive and personalized learning (ITS, adaptive hypermedia, and recommendation systems for mathematics and computer science); (3) robotics and embodied AI (KIBO, LEGO Mindstorms, botSTEM, and similar platforms in K-12 STEM); (4) generative AI and language models (ChatGPT applications, automated assessment, and AI-generated feedback, growing most rapidly post-2022); and (5) emerging technologies (VR/AR-enabled instruction, IoT-connected environments, and gamified AI platforms). These five clusters align with the four AI application categories identified in the content analysis (Section 6) and serve as the conceptual foundation for Figure 6.

FINDINGS

A comprehensive content analysis of the 54 included studies identified four primary categories of artificial intelligence (AI) applications in STEM education, as outlined by Wang et al. (2024): (1) adaptive learning and personalized tutoring (AL); (2) intelligent assessment and management (IA); (3) profiling and prediction (PP); and (4) emerging products (EP). Figure 4 illustrates the distribution of included studies across these categories. Figure 6 presents the overarching conceptual framework. A summary of findings aligned with each research question is provided in Table 7 at the conclusion of this section. Sections 6.1-6.4 delve into each AI category, while Section 6.5 synthesizes research design elements.

DOI: 10.46121/pspc.54.3.54

Power System Protection and Control
ISSN:1674-3415

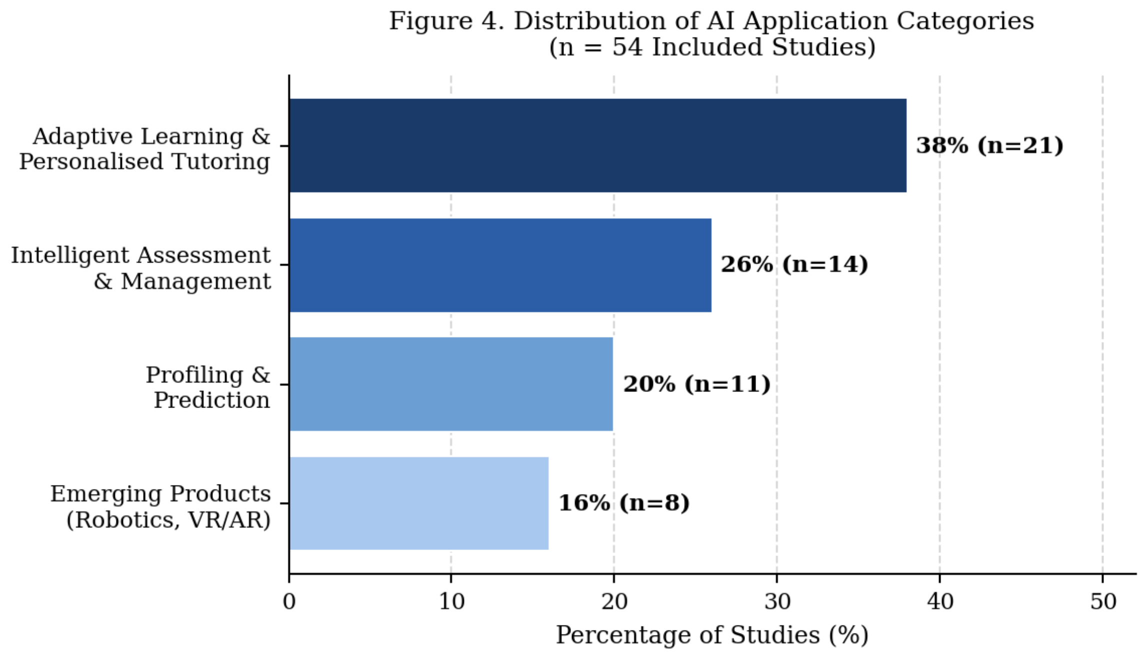

Fig. 4. Distribution of AI Application Categories Across Included Studies (n = 54)

A. Adaptive Learning and Personalised Tutoring (RQ1, RQ2)

Adaptive learning systems are the most extensively studied AI category, comprising 38% of the included studies (n=21). These applications tailor the learning process based on learners' knowledge levels, learning styles, and performance trajectories. Two prominent sub-categories within the STEM literature are intelligent tutoring systems (ITS) and adaptive learning platforms.

Intelligent tutoring systems in mathematics and computer science exhibit the most robust evidence base within the STEM domain. Bauer et al. (2025) conducted a meta-analysis that confirmed the moderate positive effects of AI-enhanced interventions on STEM achievement (d = 0.58). Ouyang et al. (2023) demonstrated that integrating AI performance prediction with adaptive feedback in engineering online courses resulted in significant improvements in outcomes when instructors were trained to interpret and respond to system outputs. This finding underscores the critical role of human-in-the-loop design as a moderating variable. Xavier et al. (2023) discovered that an AI-driven feedback tool empowered STEM instructors to personalize responses on a large scale, with the effectiveness contingent upon instructor engagement with system outputs. Rodriguez-Ortiz et al. (2025) systematically reviewed machine learning and generative AI applications for adaptive learning in computer science and engineering higher education, identifying the increasing integration of large language models (LLMs) for personalized explanations and code generation. Adaptive learning platforms extend this logic to broader curriculum content delivery. Weidlich et al. (2023) demonstrated significant gains from highly informative language-based (LA)-based feedback, but the effect was moderated by students' pre-existing feedback literacy. This finding suggests that technology-mediated personalization cannot substitute for metacognitive preparedness.

B. Intelligent Assessment and Management (RQ1, RQ2)

Intelligent assessment and management applications constitute 26% of the included studies (n=14), encompassing intelligent assessment systems (IAS) and AI-enhanced learning management systems (LMS). Intelligent assessment systems utilize AI to efficiently and consistently conduct STEM assessment tasks across large cohorts. Automated grading for mathematical proofs, programming code, and science short answers is the most advanced function. Zhao (2025) conducted a systematic review and consistently found positive evaluations of AI-assisted STEM assessment. Lee and Moore (2024) discovered that AI-generated formative feedback achieves comparable quality to instructor feedback for structured tasks while operating at a significantly larger scale. Shi and Aryadoust (2024) documented accuracy gains, particularly for grammar-level and structural features, in AI-written feedback systems applied to science writing. Li, F.N. (2023) demonstrated that AI-supported assessment integrated into curriculum redesign in sports communication education resulted in measurable course outcome improvements. AI-enhanced LMS platforms in STEM contexts facilitate collaborative learning, classroom management, and information delivery. Bridge et al. (2023) demonstrated that AI-driven personalized messaging enhanced STEM engagement, but only when students perceived messages as humanizing. Moon et al. (2023) demonstrated that learning analytics tools embedded in an LMS improved peer learning patterns in asynchronous STEM environments. Lim et al. (2023) identified alignment between learning analytics and instructional design as a prerequisite for meaningful LMS-mediated AI support.

C. Profiling and Prediction (RQ1, RQ2)

Profiling and prediction applications constitute 20% of the included studies (n=11), employing educational data mining and machine learning techniques to discern learner characteristics, anticipate learning outcomes, and identify at-risk students. Four distinct functional sub-types are delineated: failure and dropout warning; academic achievement prediction; learning analytics and learner modeling; and course arrangement support.

In 2023, Guleria and Sood presented an explainable AI career counseling framework for STEM students, achieving a 91.2% recall rate using Naive Bayes. This framework also highlighted the equity implications of access to STEM pathways. Abuzinadah et al. (2023) applied convolutional features and machine learning to predict engineering performance based on LMS activity data as early as the second week of a course. Aljawarneh et al. (2023) emphasized the importance of interpretable outputs and instructor engagement as prerequisites for translating predictions into actionable interventions in higher education STEM. Fayda-Kinik (2025) cataloged persistent barriers, including data quality, privacy, and scalability, that hinder the deployment of these AI systems in lower-resource contexts. Akinyi et al. (2023) and Ouyang et al. (2023) further demonstrated that integrating prompts and feedback into the prediction pipeline (rather than standalone alerts) yields the most effective outcomes in STEM programs.

D. Emerging Products: Robotics, Generative AI, and Immersive Technologies (RQ1, RQ2)

Emerging AI-embedded products constitute 16% of the included studies (n=8), encompassing educational robotics (including chatbots), generative AI tools, and virtual/augmented reality (VR/AR) applications. Educational robotics stands as the most theoretically grounded sub-strand, boasting a substantial K-12 evidence base. Hughes et al. (2022) demonstrated concurrent improvements in STEM skills and social competencies from the RAISE robotics program; Sisman et al. (2021) isolated gains in spatial ability and STEM attitude from robotics training; Gerosa et al. (2022) documented computational thinking gains in preschoolers; and Kabadayi and Sonmez (2024) demonstrated improvements in creative thinking in early childhood STEM. Mangina et al. (2023) conducted a scoping review of 40 robotics programs, confirming consistent computational thinking and problem-solving gains across primary and preschool contexts, with a geographic concentration in the USA and Europe. Hwang and Chang (2021) reviewed chatbot opportunities and challenges in education, identifying four chatbot functions—classroom teacher, peer support, emotional companion, and telepresence teacher—with growing STEM applications, particularly in computer science education. Generative AI tools have experienced the most rapid growth since 2022. Ali et al. (2024) found ChatGPT scored 70-100% on structured STEM knowledge items but performed poorly on image-based clinical reasoning. Herbold et al. (2023) demonstrated that ChatGPT-generated essays scored higher than human equivalents on quality metrics, raising integrity concerns that Perkins et al. (2023) confirmed cannot be adequately addressed through detection-based approaches. Kasneci et al. (2023) comprehensively reviewed the opportunities and challenges of LLMs in education, identifying personalized feedback, content generation, and automated assessment as key AI affordances. Virtual and augmented reality (VR/AR) applications exhibit strong STEM engagement effects: Li and Li (2022) documented AI-enhanced physical fitness evaluation in sports science; and Kalozy-Szabo et al. (2022) employed VR-mediated robotics to enhance cognitive development in applied STEM contexts.

E. Research Design Elements (RQ1, RQ5)

Figure 5 illustrates the distribution of research methodologies and educational settings across the 54 included studies. Experimental and quasi-experimental designs prevail (42.6%), reflecting the field's emphasis on demonstrating causal AI effects. Statistical analysis and machine learning validation account for 22.2%; survey research (12.0%), qualitative designs (11.1%), and mixed methods (12.0%) are proportionally smaller, corroborating the underrepresentation of qualitative inquiry (Wang et al., 2024).

DOI: 10.46121/pspc.54.3.54

Power System Protection and Control
ISSN:1674-3415

Figure 5. Research Methods and Educational Contexts of Included Studies

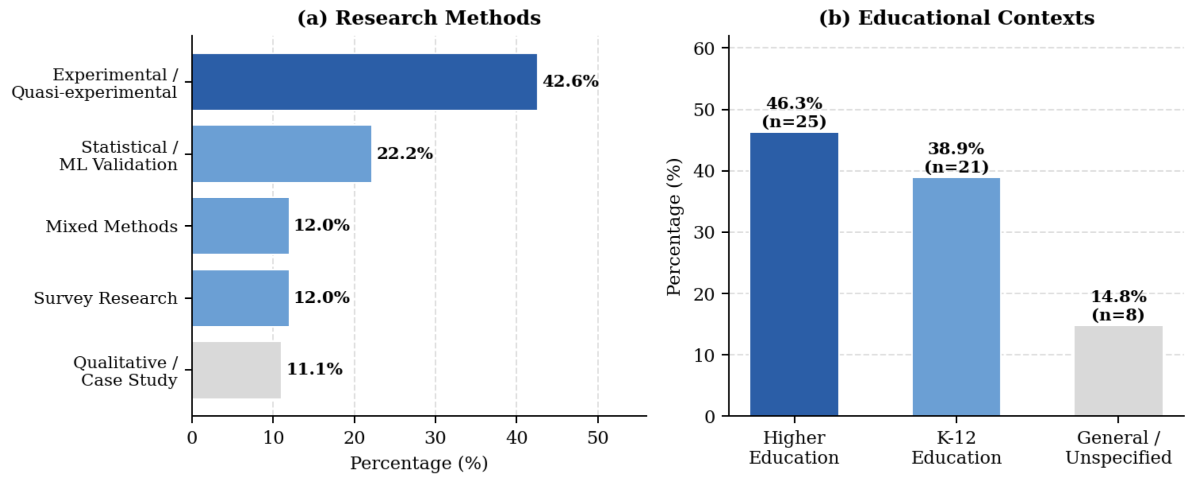

Fig. 5. Research Methods (a) and Educational Contexts (b) Across Included Studies (n = 54)

Higher education is the most extensively studied setting (46.3%), followed by K-12 education (38.9%), and unspecified settings (14.8%). The theoretical grounding of the studies varies significantly. Constructivist learning theory (n=8), cognitive load theory (n=6), computational thinking frameworks (n=5), and learning style theory (n=4) are the most frequently cited frameworks. However, 51.9% of the included studies do not explicitly invoke a theoretical framework, which is a notable structural weakness consistent with broader evidence in the field of Artificial Intelligence and Education (Wang et al., 2024). Figure 6 synthesizes the taxonomy and cross-cutting themes.

Figure 6. Conceptual Framework: AI Application Taxonomy in STEM Education

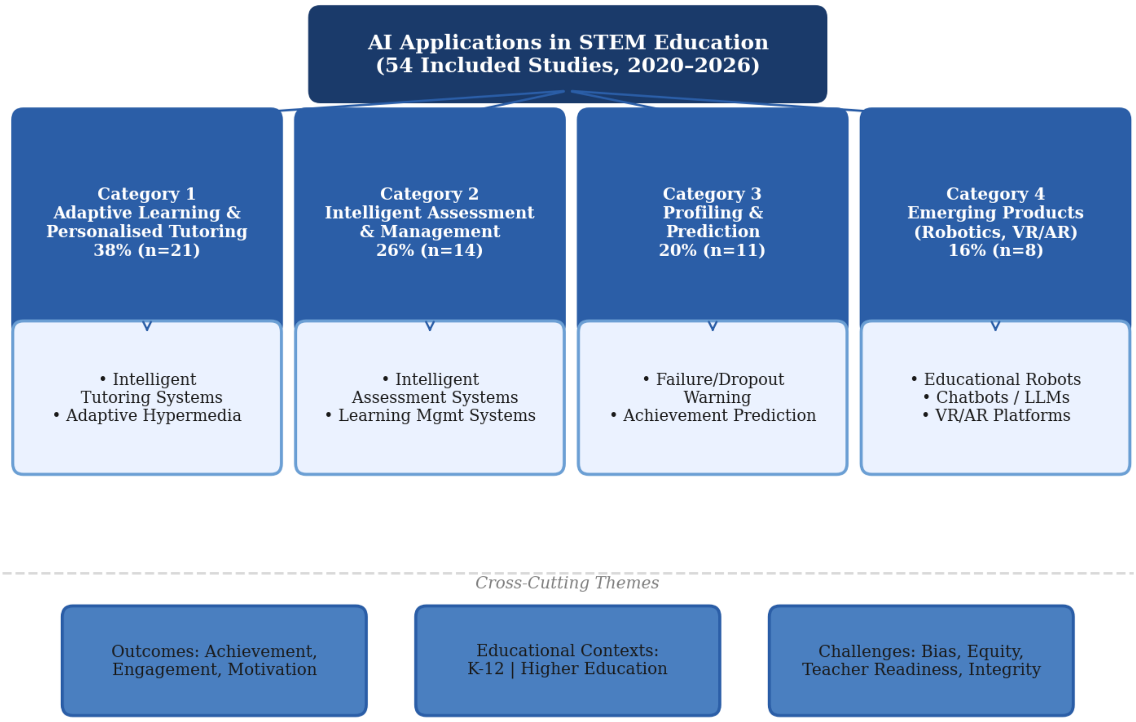

Fig. 6. Conceptual Framework: AI Application Taxonomy and Cross-Cutting Themes in STEM Education (2019-2026)

F. Summary of Findings by Research Question

TABLE VII: SUMMARY OF KEY FINDINGS MAPPED TO RESEARCH QUESTIONS

RQ	Key Finding	Representative Studies	Implication
RQ1	Four AI categories identified: AL (38%), IA (26%), PP (20%), EP (16%). Adaptive learning is most studied; emerging products growing fastest.	Ouyang et al. (2023); Hughes et al. (2022); Chan & Lee (2023)	Design frameworks must account for distinct AI categories, not treat AI as monolithic.
RQ2	AI serves four purposes: instructional scaffolding, automated assessment and feedback, teacher-side analytics, and student-directed generative AI. Positive outcomes most consistent when pedagogically aligned.	Bauer et al. (2025); Xavier et al. (2023); Weidlich et al. (2023)	Effective AI use requires pedagogical alignment, teacher training, and authentic assessment redesign.
RQ3	Top barriers: low teacher AI self-efficacy, algorithmic bias, academic integrity, access inequity. Conditions moderating success: teacher readiness, feedback literacy, institutional support.	Filiz et al. (2025); Garcia-Lopez & Trujillo-Linan (2025); Perkins et al. (2023)	Governance frameworks, bias audits, and STEM-specific professional development are urgently needed.
RQ4	Publication growth from 22.2% (2019–21) to 38.9% (2024–26). Geographic concentration in North America and East Asia. Scopus Q1/Q2 journals = 74.1%.	Wang et al. (2024); Mangina et al. (2023)	Research investment needed in Global South; multi-disciplinary STEM review needed.
RQ5	Gaps: qualitative methods underused (11.1%); 51.9% of studies lack explicit theoretical grounding; longitudinal designs absent; equity-disaggregated analyses rare.	Bauer et al. (2025); Thomas & Harden (2008)	Priority: longitudinal, equity-disaggregated, theory-driven AI-in-STEM research.

DISCUSSION

In response to RQ1 and RQ4, the bibliometric analysis reveals a rapid expansion in the publication output of AI-in-STEM fields. Geographic concentration is observed in North America and East Asia, while journal diversity is also increasing. The content analysis identifies four distinct AI application categories within the STEM field, each characterized by its unique pedagogical logic, evidence base, and implementation profile. These categories have not been previously catalogued in a STEM-specific dual-method review. Adaptive learning and personalized tutoring systems demonstrate the most robustly evidenced gains in STEM achievement ($d = 0.58$; Zhang et al., 2025), particularly in mathematics and computer science, where the knowledge structure aligns most naturally with AI-mediated instruction. Intelligent assessment systems are experiencing rapid growth and possess strong potential for providing feedback at large STEM class sizes. Profiling and prediction systems offer early-warning capabilities with significant equity implications. Emerging products, particularly robotics in K-12 education and generative AI in higher education, exhibit the fastest growth rates.

In response to RQ2, the evidence substantiates that AI technologies fulfill four educational functions within the STEM domain: instructional scaffolding, automated assessment and feedback, teacher-side analytics, and student-directed generative AI. Positive outcomes are most consistently observed when tools are tailored to a specific pedagogical requirement, teachers receive training to interpret AI outputs, and assessments are designed to reward the cognitive processes mediated by the AI rather than the products it can generate. Both learner feedback literacy (Weidlich et al., 2023) and instructor interpretation capacity (Ouyang et al., 2023; Bridge et al., 2023; Xavier et al., 2023) emerge as dominant moderating variables, indicating that human factors rather than technology specifications are the primary determinants of AI effectiveness.

In response to RQ3, three recurring barrier clusters emerge: teacher readiness (AI self-efficacy predicting adoption; Filiz et al., 2025), ethical exposure (algorithmic bias, academic integrity, differential access; Garcia-Lopez & Trujillo-Linan, 2025; Perkins et al., 2023), and institutional sustainability (access inequity

DOI: 10.46121/pspc.54.3.54

disproportionately impacting Global South STEM institutions; UNESCO, 2023). These barriers transcend technical limitations and encompass pedagogical, governance-related, and socio-political dimensions, necessitating solutions that extend beyond the design of AI systems.

In response to RQ5, the review identifies five research gaps that require immediate attention: (1) longitudinal studies to assess the long-term retention of STEM learning beyond immediate post-tests; (2) qualitative research into the implementation mechanisms of effective AI integration in STEM; (3) research in non-Western and low-resource STEM contexts, which currently constitute less than 15% of the corpus; (4) equity-disaggregated analyses as a standard practice rather than an exception in AI-in-STEM studies; and (5) theoretical work that specifies the cognitive and motivational mechanisms through which AI-mediated STEM instruction produces its effects. The underrepresentation of theoretical frameworks (51.9% of studies) constitutes a structural limitation that hinders the accumulation of generalizable knowledge and the development of evidence-based design principles for AI in STEM education.

IMPLICATIONS

For STEM educators, effective AI integration necessitates pedagogical alignment rather than mere tool adoption. The robotics literature, including botSTEM (Greca Dufranc et al., 2020), RAISE (Hughes et al., 2022), and KIBO-based curricula, exemplifies the template: explicit alignment between a theoretical framework, a pedagogical design, and a selected AI tool. This three-way alignment is rarely present in generative AI or learning analytics deployments in STEM, and its absence is the most consistent predictor of underperformance. STEM-specific professional development must transcend AI literacy to encompass disciplinary understanding of how AI tools interact with STEM epistemologies (Kim & Kim, 2022). For curriculum designers, AI should be positioned as a curricular affordance—a tool that students learn to use critically within STEM norms—rather than a content delivery add-on. Assessment redesign toward authentic, process-based STEM tasks that AI cannot easily substitute is the sustainable response to academic integrity pressures (Liang et al., 2025; Zhao, 2025; Kasneci et al., 2023). For policymakers, the equity evidence demands attention: infrastructure investment without commensurate teacher preparation, curriculum alignment, and institutional governance is insufficient. Regulatory frameworks requiring bias audits, data privacy protections, and pedagogical validation before AI tool procurement are largely absent globally (Garcia-Lopez & Trujillo-Linan, 2025) and are urgently needed. The geographic concentration of the evidence base (Figure 3) necessitates that AI-in-STEM policy in the Global South cannot rely uncritically on frameworks developed in North American and East Asian contexts. For researchers, the most urgent methodological priority is longitudinal, multi-measure designs—particularly those that follow STEM learners across full courses or years, utilize multiple types of evidence, disaggregate by equity-relevant variables, and explicitly test theorized mechanisms.

LIMITATIONS

This review presents six primary limitations. Firstly, the restriction to English-language publications excludes substantial AI-in-STEM research conducted in Chinese, Arabic, Spanish, and Portuguese. Secondly, publication bias in the underlying literature likely overstates positive AI effects, as negative or null-finding studies are underrepresented across all included databases. Thirdly, the bibliometric analysis is based on 54 included studies rather than the full 1,267-record corpus, limiting the generalizability of publication trend and conceptual cluster findings. A comprehensive corpus-level bibliometric analysis is recommended for a follow-up study. Fourthly, the high heterogeneity of technology types, STEM disciplines, educational levels, and outcome measures prevents meta-analytic pooling for most sub-questions. Thematic synthesis identifies patterns without establishing causality. Fifthly, the rapid pace of AI development means that pre-2023 tool evaluations may already describe a superseded landscape. Future reviews should consider living systematic review protocols (Elliott et al., 2017). Sixthly, while inter-rater reliability was satisfactory (screening kappa = 0.82; full-text kappa = 0.79; quality ICC = 0.78), the review was conducted by a small team without an external advisory panel, which may introduce perspective biases in coding decisions.

CONCLUSION

This systematic literature review synthesizes evidence from 54 empirical studies on AI technologies in STEM education from 2019 to 2026. It combines PRISMA 2020-guided selection, MMAT quality appraisal, bibliometric analysis, and thematic synthesis following Carrera-Rivera et al. (2022) and Wang et al. (2024). Four AI application

DOI: 10.46121/pspc.54.3.54

categories are identified: adaptive learning and personalized tutoring, intelligent assessment and management, profiling and prediction, and emerging products. Each category has distinct pedagogical logics, evidence profiles, and implementation conditions. The publication volume has grown sharply (Figure 2), geographically concentrated in North America and East Asia (Figure 3). Five conceptual keyword clusters map onto the application taxonomy. Experimental methods dominate (Figure 5a), and more than half of studies lack explicit theoretical grounding. The conceptual framework (Figure 6) and RQ summary (Table 7) organize the evidence for future researchers and practitioners.

AI benefits STEM learners most consistently when: (1) technology selection is driven by a specific, understood pedagogical need; (2) teachers possess both disciplinary and AI-specific knowledge to interpret and act on system outputs; (3) assessments are redesigned for authentic STEM performance; and (4) institutional governance actively manages algorithmic bias, data privacy, and equitable access. The proliferation of AI in STEM education is accelerating faster than the research community can evaluate it, teacher education systems can prepare for it, and policy communities can govern it. Closing these empirical, professional, and regulatory gaps—prioritizing equity and disciplinary specificity at every step—is the defining work of the coming decade in STEM education research.

DECLARATIONS

Funding: This research received no external funding.

Data Availability Statement: The full extraction data supporting the findings of this review are available from the corresponding author upon reasonable request.

Competing Interests: The author declares no competing interests.

Ethics Statement: This systematic review did not involve primary data collection from human participants; therefore, no ethics approval was required.

REFERENCES

1. Abuzinadah, N., Farouk, M., Alharthi, M., Sadad, T., Hussain, A., & Elmedany, W. (2023). Role of convolutional features and machine learning for predicting student academic performance from MOODLE data. PLoS One, 18(11), e0292131.
2. Akinyi, G. L., Oboko, R., & Muchemi, L. (2023). Learning analytics intervention using prompts and feedback for measurement of e-learners socially-shared regulated learning. Electronic Journal of E-Learning, 21(4), 111–122.
3. Ali, K., Barhom, N., Tamimi, F., & Duggal, M. (2024). ChatGPT - A double-edged sword for healthcare education? Implications for assessments of dental students. European Journal of Dental Education, 28(1), 206–211.
4. Aljawarneh, S., Lara, J. A., Vaquero-Gonzalez, L., & Atif, A. (2023). Using predictive learning analytics to support students in higher education. Education and Information Technologies, 28(8), 9871–9889.
5. Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. Educational Psychology Review, 37(2). https://doi.org/10.1007/s10648-025-10059-6
6. Biri, S. K., Kumar, S., Panigrahi, M., Mondal, S., Behera, J. K., & Mondal, H. (2023). Assessing the utilization of large language models in medical education. Cureus. https://doi.org/10.7759/cureus.47468

7. Bridge, C., Cohen, J., & Liu, D. Y. T. (2023). "I feel seen": What higher education students find helpful in learning analytics-informed personalised messaging from teachers. Journal of University Teaching and Learning Practice, 20(3).
8. Carrera-Rivera, A., Ochoa, W., Larrea-Araujo, C., & Ayala-Chauvin, M. (2022). How to conduct a systematic literature review: A quick guide for computer science research. MethodsX, 9, 101736. https://doi.org/10.1016/j.mex.2022.101736
9. Celik, I., Dindar, M., Muukkonen, H., & Jarvela, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review. TechTrends, 66(4), 616–630.
10. Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI than their Gen X and millennial teachers? Smart Learning Environments, 10(1), 60.
11. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20(1), 22.
12. Dadson, Y. A., Mohammed, Z., Hussein, B., Armah, Y. E., & Kumi-Yeboah, A. (2025). Ethics and privacy in AI education. In Future Education and Learning Spaces (pp. 215–242). Springer.
13. Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., & Thomas, J. (2017). Living systematic review: 1. Introduction — The why, what, when, and how. Journal of Clinical Epidemiology, 91, 23–30.
14. Fayda-Kinik, H. (2025). Challenges in implementing predictive analytics for personalized learning. Journal of Learning Analytics, 12(1), 45–62.
15. Filiz, O., Kaya, M. H., & Adiguzel, T. (2025). Teachers and AI: Understanding the factors influencing AI integration in K-12 education. Education and Information Technologies, 30(13), 17931–17967.
16. Garcia-Lopez, I. M., & Trujillo-Linan, L. (2025). Ethical and regulatory challenges of generative AI in education: A systematic review. Frontiers in Education, 10, 1565938.
17. Garzón, J., Patino, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. Multimodal Technologies and Interaction, 9(8), 84.
18. Gerosa, A., Koleszar, V., Tejera, G., Gomez-Sena, L., & Carboni, A. (2022). Educational robotics intervention to foster computational thinking in preschoolers. Frontiers in Education, 7, 824234.
19. Greca Dufranc, I. M., Garcia Terceno, E., Fridberg, M., Cronquist, B., & Redfors, A. (2020). Robotics and early-years STEM education: The botSTEM framework and activities. European Journal of STEM Education, 5(1), 01.
20. Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: Performance evaluation on educational data mining inspired career counseling. Education and Information Technologies, 28(1), 1081–1116.
21. Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. Scientific Reports, 13(1), 18617.
22. Hong, Q. N., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., & Pluye, P. (2018). The Mixed Methods Appraisal Tool

(MMAT) version 2018 for information professionals and researchers. Education for Information, 34(4), 285–291.

23. Hughes, C. E., Dieker, L. A., Glavey, E. M., Hines, R. A., Wilkins, I., Ingraham, K., & Taylor, M. S. (2022). RAISE: Robotics & AI to improve STEM and social skills for elementary school students. Journal of Research in STEM Education, 8(1), 1–22.
24. Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. Interactive Learning Environments, 29(7), 1077–1088.
25. Ifenthaler, D., Schumacher, C., & Kuzilek, J. (2023). Investigating students use of self-assessments in higher education using learning analytics. Journal of Computer Assisted Learning, 39(1), 255–268.
26. Indran, I. R., Paranthaman, P., Gupta, N., & Mustafa, N. (2024). Twelve tips to leverage AI for efficient medical question generation. Medical Teacher, 46(8), 1021–1026.
27. Jin, F. J. Y., Yan, L. X., Martinez-Maldonado, R., Gasevic, D., Chan, P. W. K., & Tsai, Y. S. (2023). Latent classes of self-reported feedback experiences. Computers in Human Behavior, 148, 107886.
28. Kabadayi, G., & Sonmez, D. (2024). The effects of a robotics program on preschool-students creative thinking skills. Hacettepe University Journal of Education, 39(2), 1–15.
29. Kalozy-Szabo, C., Mohai, K., & Cottini, M. (2022). Employing robotics in education to enhance cognitive development. Sustainability, 14(23), 15951.
30. Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nemmert, C., Nie, G., Rottger, L., Schuhmann, C., Sina, L., Steinle, S., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.
31. Kazakoff, E. R., Sullivan, A., & Bers, M. U. (2013). The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. Early Childhood Education Journal, 41(4), 245–255.
32. Khan, R., & Habib, M. (2025). Bridging feedback and prediction: A review of learning analytics in education. International Journal of Educational Technology, 22(1), 89–108.
33. Kim, N. J., & Kim, M. K. (2022). Teachers perceptions of using an AI-based educational tool for scientific writing. Frontiers in Education, 7, 755914.
34. Lamptey, D. L., Cagliostro, E., Srikanthan, D., Hong, S., Dief, S., & Lindsay, S. (2021). Assessing the impact of an adapted robotics programme on STEM interest among children with disabilities. Disability and Rehabilitation: Assistive Technology, 16(5), 551–558.
35. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.
36. Lee, A. V. Y., Tan, S. C., & Teo, C. L. (2023). Designs and practices using generative AI for sustainable student discourse and knowledge creation. Smart Learning Environments, 10(1), 59.
37. Lee, J., Yunus, S. A., & Lee, J. O. (2025). Investigating children programming skills through play with robots (KIBO). Early Childhood Education Journal, 53(1), 109–117.

38. Lee, S. S., & Moore, R. L. (2024). Harnessing generative AI for automated feedback in higher education: A systematic review. Online Learning, 28(3).
39. Lewis, S., Heath, G., Lim, L., & Roberts, R. (2023). Supporting commencing university students through technology-mediated personalised communication. Student Success, 14(2), 1–12.
40. Li, F. N. (2023). AI-enhanced curriculum design and deep-learning-based assessment in international sports communication education. International Journal of Information and Communication Technology Education, 19(1), 1–16.
41. Li, Y., & Li, X. (2022). The artificial intelligence system for sports education guidance and physical fitness evaluation under deep learning. Frontiers in Public Health, 10, 917053.
42. Liang, J., Stephens, J. M., & Brown, G. T. L. (2025). A systematic review of the early impact of AI on higher education curriculum, instruction, and assessment. Frontiers in Education, 10.
43. Lim, L. A., Atif, A., Heggart, K., & Sutton, N. (2023). In search of alignment between learning analytics and learning design. Education Sciences, 13(4), 411.
44. Lindsay, S., Kolne, K., Oh, A., & Cagliostro, E. (2019). Children with disabilities engaging in STEM: A group-based robotics program. Canadian Journal of Science, Mathematics and Technology Education, 19(3), 248–264.
45. Liu, J., Wang, L., & Zhou, H. (2021). Human-computer interaction fused with artificial intelligence in sports moving target detection education for college athletes. Frontiers in Psychology, 12, 677590.
46. Mangina, E., Psyrra, G., Screpanti, L., & Scaradozzi, D. (2023). Robotics in primary and preschool education: A scoping review. IEEE Transactions on Learning Technologies, 17, 342–363.
47. Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. Journal of Experimental Child Psychology, 160, 92–106.
48. [48] Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. Education Sciences, 13(9), 856.
49. Moon, J., McNeill, L., Edmonds, C. T., Banihashem, S. K., & Noroozi, O. (2023). Using learning analytics to explore peer learning patterns in asynchronous gamified environments. International Journal of Emerging Technologies in Learning, 18(11).
50. Ogbuchi, I. (2023). Leveraging machine learning and AI to enhance educational learning analytics. International Journal on Studies in Education, 5(1), 65–81.
51. Ogurlu, U., & Mossholder, J. (2023). The perception of ChatGPT among educators: Preliminary findings. Research in Social Sciences and Technology, 8(4), 196–215.
52. Ouyang, F., Wu, M., Zheng, L. Y., Zhang, L. Y., & Jiao, P. C. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering courses. npj Science of Learning, 8(1), 1.
53. 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., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... & Moher,

DOI: 10.46121/pspc.54.3.54

D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.
54. Papamitsiou, Z., & Economides, A. A. (2022). Learning analytics and feedback: A review of current research. Computers & Education, 184, 104529.
55. Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2023). Detection of GPT-4 generated text in higher education. Journal of Academic Ethics. https://doi.org/10.1007/s10805-023-09492-6
56. Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Blackwell.
57. Rodriguez-Ortiz, M. A., Santana-Mancilla, P. C., & Anido-Rifon, L. E. (2025). Machine learning and generative AI in learning analytics for higher education. Applied Sciences, 15(6), 3012.
58. Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. ReCALL, 36(2), 187–209.
59. Shi, L., & Choi, I. (2024). A systematic review on artificial intelligence in supporting teaching practice. In AI-based STEM Instruction and Teacher Professional Development (pp. 321–347). Oxford University Press.
60. Sisman, B., Kucuk, S., & Yaman, Y. (2021). The effects of robotics training on children spatial ability and attitude toward STEM. International Journal of Social Robotics, 13(2), 379–389.
61. Sullivan, A., & Bers, M. U. (2016). Robotics in the early childhood classroom: Learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. International Journal of Technology and Design Education, 26(1), 3–20.
62. Sullivan, A., & Bers, M. U. (2019). Investigating the use of robotics to increase girls interest in engineering during early elementary school. International Journal of Technology and Design Education, 29(5), 1033–1051.
63. Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8(1), 45.
64. Tselegkaridis, S., & Sapounidis, T. (2022). Exploring the features of educational robotics and STEM research in primary education. Education Sciences, 12(5), 305.
65. UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/ewzm9535
66. UNESCO. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/zjte2084
67. Van Den Berg, G., & Du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for lesson planning, critical thinking and openness in teacher education. Education Sciences, 13(10), 998.
68. Wang, S., Liu, H., & Zhang, X. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, 124167.
69. Watermeyer, R., Phipps, L., Lanclos, D., & Knight, C. (2024). Generative AI and the automating of academia. Postdigital Science and Education, 6(2), 446–466.
70. Weidlich, J., Fink, A., Frey, A., Jivet, I., Gombert, S., Menzel, L., Giorgashvili, T., Yau, J., & Drachsler, H. (2023). Highly informative feedback using learning analytics: How feedback literacy moderates learning. Computers in Human Behavior, 148, 107836.

71. Xavier, C., Rodrigues, L., Costa, N., Neto, R., Alves, G., Falcao, T. P., Gasevic, D., & Mello, R. F. (2023). Empowering instructors with AI: Evaluating the impact of an AI-driven feedback tool in learning analytics. Education and Information Technologies, 28, 13499–13520.
72. Younas, M., Ismayil, I., El-Dakhs, D. A. S., & Anwar, B. (2025). Exploring the impact of artificial intelligence in advancing smart learning in education: A meta-analysis. Open Praxis, 17(3), 59.
73. Zawacki-Richter, O., Marin, 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(1), 39.
74. Zhang, J., Jantakoon, T., & Laoha, R. (2025). Meta-analysis of artificial intelligence in education. Higher Education Studies, 15(2), 189.
75. Zhao, C. (2025). AI-assisted assessment in higher education: A systematic review. Journal of Educational Technology and Innovation, 6(4).