AI-Mediated STEM Problem Solving: Activity Distribution and Functional Roles in a Systematic Review

Mika Lim
University Innovation Center, Seoul National University, Seoul, Republic of Korea

***Abstract*—Artificial intelligence (AI) is increasingly embedded in STEM problem-solving activities, yet existing reviews have generally organized the literature by technology type or learning outcome. This review shifts the analytic focus from what AI can do to how it is integrated into students' STEM problem solving. It examined where AI was positioned within problem-solving activity, what functional role it performed, and how its functions or outputs were connected to students' subsequent judgments and actions. Following PRISMA 2020, searches of Scopus, Web of Science, ERIC, and backward references identified 29 English-language journal articles published from 2017 through June 4, 2026. Study-level coding mapped AI use across seven non-linear activity categories, while case-level analysis examined 37 reporting-eligible AI–student interactions in which the AI function or output, activity context, and subsequent student action were identifiable. AI was reported most often in two activity categories: Problem Understanding and Exploration, and Revision and Improvement. Four recurring functional roles were identified: Conceptual Resource, Generative Resource, Operational Component, and Adaptive Scaffold. The same technology assumed different roles depending on the activity category and how students used its output. The principal contribution of this review is to shift attention from AI's technical capabilities and associated learning outcomes to the interaction between AI outputs and students' subsequent activity. By linking each AI function or output to a problem-solving activity category and a subsequent student action, the review shows that AI's role is defined by how students use its output in context, not by the technology itself. The findings are descriptive of the selected corpus and its reporting practices; they should not be interpreted as prevalence estimates or evidence of causal learning effects.**

***Keywords*—Artificial intelligence; STEM education; AI-mediated learning; human-AI interaction; functional roles; systematic review**

I. Introduction

A. Problem and Purpose

Artificial intelligence (AI) has become a prominent component of educational technology. Its educational uses now extend beyond learner diagnosis, performance prediction, and intelligent tutoring to explanations of concepts, generation of ideas and code, model execution, data analysis, feedback, and support for reflection. Generative AI and large language models (LLMs), in particular, enable students to ask follow-up questions, explore alternative solutions, review emerging artifacts, and identify possible directions for revision [1]–[3].

Introducing AI into a classroom does not by itself establish educational value. The same system may serve as an information source for conceptual understanding, generate design alternatives, operate as part of a physical device or computational model, or diagnose a student-generated artifact and suggest next steps. Analysis must therefore consider not only what an AI system can do but also where its output enters students' activity and how students use that output in subsequent judgments and actions.

Within science, technology, engineering, and mathematics (STEM) education, students act as investigators and designers who coordinate scientific concepts, mathematical reasoning, engineering design, and technological tools in addressing authentic problems [4], [5]. They identify the problem and its requirements, seek relevant knowledge and evidence, formulate and implement a solution, examine results through execution or experimentation, interpret evidence, and revise the solution. Throughout this process, they form and revise judgments while engaging with data, tools, phenomena, and artifacts.

AI can be integrated at multiple points in this process. Intelligent tutoring systems can provide explanations and hints based on students' current understanding. Generative AI can answer questions and generate ideas, designs, code, and explanatory drafts. Educational robots, sensors, and artificial intelligence of things (AIoT) systems can sense conditions, collect data, execute models, and control devices. These technical functions are not equivalent to functional roles. Code generation, object recognition, data analysis, and automated feedback describe technical operations; their functional roles depend on how students use the resulting outputs.

Previous systematic reviews have primarily organized research on AI in education by technology type, system function, research method, or learning outcome. Zawacki-Richter et al. classified higher-education applications into profiling and prediction, assessment, adaptive systems, and intelligent tutoring [6]. Xu and Ouyang analyzed AI-STEM research in terms of technical applications and components of educational systems [3]. Akhmetova et al. examined technologies, methods, and cognitive and affective outcomes in high-school STEM education [1], whereas Pacheco-Olea and Villacís-Macías synthesized study characteristics, technologies, outcomes, and methodological limitations across AI-STEM research [2].

These reviews clarify the scope and direction of the field, but technology-centered classifications do not fully explain how the same AI technology can perform different roles across problem-solving activities. Outcome-centered syntheses can identify differences associated with AI use but often do not specify the activities in which those differences emerged. The present review therefore examined a defined corpus of empirical AI-STEM journal articles to identify where AI was reported within students' problem-solving activity and which recurring functional roles were visible in well-documented AI–student interactions. The study was designed as a descriptive synthesis of reported activity and interaction patterns, not as an estimate of prevalence or evidence that AI improved learning. Its purpose was to inform more precise descriptions and evaluations of AI-supported STEM learning environments.

B. Research Questions

RQ1. In which activity categories of students' STEM problem solving was AI reported?

RQ2. What recurring functional roles were reported in AI-mediated STEM problem-solving interactions?

II. Theoretical Background

A. Activity Structure of STEM Problem Solving

STEM education integrates scientific inquiry, engineering design, mathematical thinking, and technology use [4]. Accordingly, learning extends beyond the acquisition of disciplinary knowledge to the exploration of relevant concepts and evidence, the use of tools, and the design, implementation, and testing of solutions to complex problems.

Comparable activity structures appear in science and engineering practices, engineering design, problem-based learning, and project-based learning. Science and engineering practices emphasize asking questions and defining problems, developing models, conducting investigations, analyzing data, constructing explanations, and designing solutions [5], [7]. Engineering design emphasizes identifying requirements and constraints, generating alternatives, implementing and testing solutions, making decisions, and improving solutions iteratively [8]. Problem-based and project-based learning organize inquiry, artifact construction, and review around authentic problems [9], [10]. Although these approaches differ in terminology and emphasis, they share a progression from initial engagement with a problem through planning and action to testing, interpretation, revision, and reflection.

On this basis, the present analysis distinguished seven activity categories: Problem Understanding and Exploration, Planning and Design, Implementation and Construction, Experimentation and Testing, Analysis and Interpretation, Revision and Improvement, and Reflection. The categories are not proposed as a new general model of STEM problem solving. Rather, they provide a common, non-linear coding frame for comparing studies grounded in science and engineering practices, engineering design, problem-based learning, and project-based learning. Table I presents the operational definition of each category.

TABLE I. Operational Definitions of STEM Problem-Solving Activity Categories

Activity Category	Operational Definition
A1. Problem Understanding and Exploration	Identifying the problem situation and requirements and exploring relevant concepts, examples, and information.
A2. Planning and Design	Generating alternatives and specifying a strategy, procedure, or design while considering conditions and constraints.
A3. Implementation and Construction	Producing an artifact such as code, a model, a device, a program, a product, or an explanation.
A4. Experimentation and Testing	Executing, observing, or measuring an artifact or system to examine operation and performance.
A5. Analysis and Interpretation	Identifying errors, patterns, relationships, and meaning in collected data and execution results.
A6. Revision and Improvement	Modifying an idea or artifact and trying again on the basis of analysis or feedback.
A7. Reflection	Explaining the problem-solving process and grounds for judgment and metacognitively examining strategies and outcomes.

Students may move among these categories rather than follow a fixed sequence. Analytically, the framework separates planning from construction, testing from learner interpretation, and artifact revision from metacognitive reflection. These distinctions matter in AI-mediated activity because a system may generate, execute, or diagnose without evidence that students interpreted its output or reflected on the grounds for a decision.

B. Technical Functions and Functional Roles of AI

In STEM education, AI supports information seeking, conceptual understanding, design, construction, experimentation, analysis, and feedback. Intelligent tutors and adaptive systems diagnose learner responses and provide explanations, hints, and learning paths. AI-based learning analytics detect patterns in performance data and present predictions or visualizations. Computer vision, machine learning, educational robotics, and AIoT combine sensing, classification, measurement, prediction, and control with experimentation and construction. Generative AI and LLMs explain concepts, answer questions, generate ideas and design alternatives, produce code, and provide error diagnoses and feedback.

A functional role cannot be inferred from a technology label or technical function alone. An LLM may provide explanations and examples during Problem Understanding and Exploration, generate alternatives during Planning and Design, draft code during Implementation and Construction, or diagnose an existing student artifact during Revision and Improvement. Computer vision and AIoT may automate a device while also making phenomena that are otherwise difficult to observe available for experimentation and interpretation.

This distinction requires five related constructs to be kept separate. A technical configuration describes the system or combination of models, interfaces, sensors, and devices. An AI function is the operation performed, such as generation, classification, prediction, or control. An AI output is the information, candidate solution, inference result, or feedback made available by that operation. A functional role describes how the function or output is incorporated into a student's problem-solving activity. A subsequent student action is the observable response that follows, such as asking a follow-up question, selecting or revising a design, testing a model, adding data, or reconsidering a justification.

Prior reviews classify research on AI by technology, system function, educational level, subject area, or outcome [1]–[3], [6]. Such classifications map the field but do not fully explain how an AI function or output becomes connected to students' problem-solving actions. Automated error detection, for example, does not establish that a student interpreted the error or revised a solution. The present review therefore uses mediation conservatively to denote a reported connection among an AI function or output, a specific STEM problem-solving activity, and an observable subsequent student action. It does not imply that AI use caused improvement in learning outcomes.

III. METHODS

A. Research Design

This systematic review identified and selected empirical studies of AI use in STEM education and analyzed both where AI was positioned in students' problem-solving activities and what functional roles it performed. Search, selection, and reporting procedures followed PRISMA 2020 guidance [11]. The analysis combined study-level descriptive quantitative analysis with case-level qualitative content analysis [14].

For clarity, study refers to an included journal article at the document level, whereas case refers to a clearly bounded AI–student interaction reported within a study. For RQ1, each included study was coded for AI use across the seven STEM problem-solving activity categories. For RQ2, the unit of analysis was an AI–student interaction case. Each case linked an identifiable AI function or output to the student activity in which AI was used and to a subsequent student action.

B. Search Strategy

The review protocol was limited to English-language journal articles indexed in Scopus, Web of Science, and ERIC and published from 2017 through June 4, 2026. The search combined an AI-related term set with a STEM-education term set, as shown in Table II.

TABLE II. DATABASE SEARCH RESULTS

Field	Details
Database records	Scopus: 320; Web of Science: 257; ERIC: 241; total: 818.
Common search string	("artificial intelligence" OR "generative AI" OR ChatGPT OR "large language model*" OR "intelligent tutoring system*" OR "educational robotics") AND ("STEM education" OR "STEM learning" OR "STEM teaching")
Period and conditions	2017–June 4, 2026; English-language journal articles indexed in Scopus, Web of Science, or ERIC

For this review, AI-supported STEM education was limited to contexts in which students directly used AI during problem exploration, design, implementation, experimentation, data interpretation, or iterative improvement. Studies were excluded when AI was used only to provide theoretical explanations or routine problem solving in a single subject, and no direct connection to STEM problem-solving activity could be identified. Table III presents the inclusion and exclusion criteria.

TABLE III. INCLUSION AND EXCLUSION CRITERIA

Criterion	Definition
Inclusion	AI technology was applied in a STEM teaching or learning context; learner activity, instructional design, an educational program, or a learning environment was clearly described; and an AI function or output could be connected to a student's STEM activity.
Exclusion	AI was not applied in an actual lesson or learning process; the context was not STEM education; AI was only a learning topic, and no STEM practice or student activity was clear; the study reported only perceptions, attitudes, satisfaction, or intention to use; or AI was used only for teacher support or system administration without a clear connection to student activity.

C. Study Selection

The database search identified 818 records. After 270 duplicates were removed on the basis of DOI and title, 548 records remained for title-and-abstract screening. Two coders participated in study selection. Coder 1 held a doctoral degree in technology education. Coder 2 was a science education teacher with more than 20 years of experience.

Coder 1 screened all 548 titles and abstracts, and Coder 2 independently screened a random sample of 110 records (20.1%). The coders agreed on 99 of the 110 decisions, yielding 90.0% agreement. This stage excluded 413 records and retained 135 articles for full-text assessment.

Both coders independently assessed all 135 full-text articles. They reached the same inclusion or exclusion decision for 125 articles, yielding 92.6% agreement and a Cohen's kappa of 0.79 [12]. For the 10 disagreements, the coders compared the full texts with the inclusion and exclusion criteria and reached consensus. The database-search pathway yielded 25 included studies.

Backward reference searching was then conducted on the reference lists of the 25 included studies, following Wohlin [13]. This search identified 1,362 reference records. After duplicates were removed and the publication-year criterion was applied, 991 records were screened by title and abstract. Twenty-three articles proceeded to full-text assessment. The two coders independently judged eligibility, with 91.3% agreement and a Cohen's kappa of 0.70 [12]. Disagreements were resolved through discussion. Four additional studies met the criteria, yielding a final corpus of 29 studies [15]–[43]. Fig. 1 summarizes the selection process.

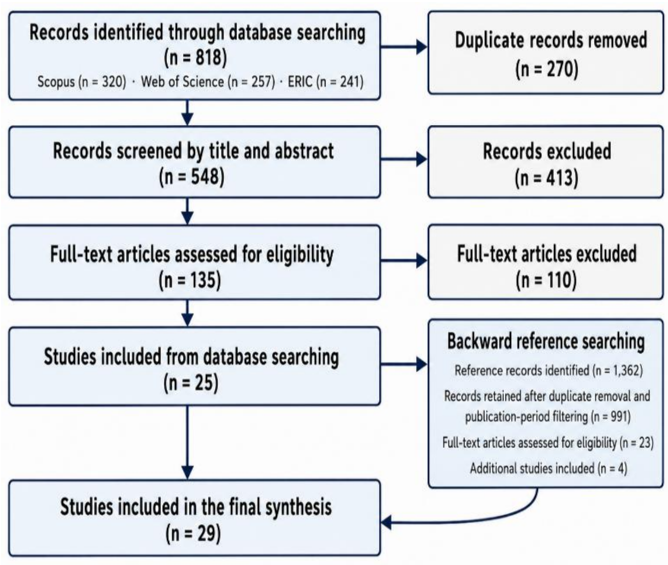

Fig. 1. Flow diagram of study identification and selection.

D. Data Extraction and Coding

Extracted data included publication year, educational level, representative technical configuration, the STEM problem-solving activity categories in which AI was directly used, and case-level information on AI–student interaction and reporting eligibility. Table IV summarizes the extraction and coding framework.

TABLE IV. DATA EXTRACTION AND CODING OVERVIEW

Level	Analytical Area	Extracted or Coded Information
Study level	Basic characteristics	Publication year, educational level, and representative technical configuration.
	STEM activity category	Problem Understanding and Exploration; Planning and Design; Implementation and Construction; Experimentation and Testing; Analysis and Interpretation; Revision and Improvement; Reflection.
Case level	AI-use case	What AI performed or provided; the student activity in which AI was used; and the reported subsequent student action.
	Functional role	Non-mutually exclusive role assigned from the relationship among AI function or output, activity context, and subsequent student action.

a) Study-level coding of activity categories: An activity category was coded only when an AI output or operation was connected to a student's STEM activity. Multiple categories could be coded within a single study. Internal system analysis alone did not justify coding Analysis and Interpretation; students had to examine an analytical or diagnostic output and construct meaning from it. General satisfaction or post-activity impressions were not coded as Reflection. Reflection required evidence that an AI-generated question or feedback was used for self-assessment, examination of the grounds for a judgment, or reconstruction of an explanation.

The two coders independently made 203 binary coding decisions (29 studies across seven categories). They agreed on 187 decisions, yielding 92.1% overall agreement and a pooled Cohen's kappa of 0.812 [12]; category-specific agreement ranged from 79.3% to 100.0%. Agreement was lowest for Analysis and Interpretation. The coders therefore revisited the boundary between internal system analysis and students' use of an analytical output. All disagreements were resolved by comparing the source text with the coding criteria, and the reported counts were calculated from the final consensus codes.

b) Case-level extraction and derivation of functional roles: For RQ2, the unit of analysis was a bounded AI–student interaction rather than an entire study. A case was retained for role analysis only when the source article explicitly documented three elements: an identifiable AI function or output, the STEM problem-solving activity in which it was used, and a subsequent student action linked to that output. Repeated instances of the same function and student action within one task were consolidated, whereas a change in the AI output, activity context, or subsequent action defined a separate case. Applying this conservative rule yielded 37 reporting-eligible cases. Interactions lacking one or more of the three elements were excluded from role-frequency analysis.

Functional roles were assigned at the interaction-case level, not to a technology, operation, or output in isolation. The AI operation and output were recorded as descriptive evidence; the role label captured how the output entered the student's problem-solving activity, as shown by the subsequent action. Both coders first summarized the three case elements and developed descriptive codes closely grounded in the source text without applying the final role labels. They then compared codes across cases, consolidated related patterns, refined inclusion and boundary rules against the source articles, and reapplied the provisional categories. After the codebook was finalized, the four non-mutually exclusive roles were represented as 148 binary decisions across the 37 cases. The coders agreed on 130 of the 148 decisions (87.8%), yielding a pooled Cohen's kappa of 0.723 [12]. Final role assignments were non-mutually exclusive: a case could receive more than one role when distinct output-action relationships were documented. The resulting framework is therefore a provisional classification of the 37 reporting-eligible cases rather than a prevalence estimate or a universal taxonomy. Table V presents the codebook used to distinguish the four functional roles.

TABLE V. CODEBOOK FOR AI FUNCTIONAL ROLES

Functional Role	AI Function or Output	Required Student Use or Subsequent Action	Exclusion or Boundary
F1. Conceptual Resource	Explanation, example, question-and-answer exchange, contextual information, or representational support.	The student uses the output to clarify a problem or concept, refine a problem representation, ask a follow-up question, or apply an explanation.	Excludes candidate solutions and feedback contingent on a specific student artifact or performance state.
F2. Generative Resource	Candidate idea, strategy, plan, design, code, model, or draft.	The student compares, selects, modifies, negotiates, implements, or rejects the candidate.	Excludes concept explanation and task-embedded AI execution.
F3. Operational Component	Sensing, detection, classification, inference, measurement, simulation, or control that produces a task-relevant output.	The student integrates, observes, tests, or interprets the operation or its performance.	Excludes internal processing that only triggers feedback and yields no task output available to the student.
F4. Adaptive Scaffold	Diagnosis of a specific student response, artifact, performance state, or task state, followed by contingent guidance.	The student revises, retries, seeks additional evidence, or reconsiders a judgment in response to feedback, hints, revision directions, or reflective questions.	Excludes general encouragement and explanation not adapted to evidence about the learner, artifact, or task state.

IV. RESULTS

A. Characteristics of the Included Studies

The final corpus comprised 29 studies that reported students using AI during STEM problem-solving activities. Sixteen studies were published in 2025 and seven by June 4, 2026; together, these 23 studies accounted for 79.3% of the corpus. The findings therefore describe a recent cross-section of a rapidly developing literature rather than a longitudinal account of change across the full 2017-2026 period.

Higher education accounted for 18 studies, secondary education for eight, primary or other child learners for two, and early childhood education for one. Table VI summarizes the selected study corpus.

TABLE VI. CHARACTERISTICS OF THE SELECTED STUDY CORPUS

Characteristic	Category	Studies (n)	Percent (%)
Publication year	2020	1	3.4
	2021	3	10.3
	2023	1	3.4
	2024	1	3.4
	2025	16	55.2
	2026	7	24.1
Educational level	Early childhood	1	3.4
	Primary or other child learners	2	6.9
	Secondary education	8	27.6
	Higher education	18	62.1

B. AI Technologies and Technical Configurations

The technologies were grouped into three representative configurations according to the system that engaged most directly and continuously with student activity. The categories were mutually exclusive at the study level and described the predominant configuration; they were not intended to classify every technology reported in each study.

Generative-AI-centered conversational and agent-based systems were the predominant configuration in 17 of 29 studies. This group included direct use of general-purpose systems such as ChatGPT and Gemini, as well as specialized systems that combined LLMs with multiple agents, structured prompting, retrieval-augmented generation, semantic routing, speech interfaces, or discipline-specific knowledge bases.

Computer-vision- and machine-learning-centered recognition and execution systems were the predominant configuration in 7 of 29 studies. Image classification, object detection, state recognition, and performance prediction were combined with JetBot, Jetson Nano, ESP32-CAM, sensors, motors, or AIoT devices. In some studies, students collected training data, retrained models, or examined misclassifications to investigate system limitations and operating conditions.

Nongenerative intelligent tutors or adaptive support systems were the predominant configuration in 5 of 29 studies. These systems analyzed student responses, performance states, or learning histories and supplied explanations, hints, task recommendations, or learning paths. Table VII summarizes the configurations.

TABLE VII. REPRESENTATIVE AI TECHNICAL CONFIGURATIONS

Representative Configuration	Examples of Technologies and Systems	Principal Technical Features	Studies (n/29)	Percent (%)
Generative-AI-centered conversational and agent-based systems	ChatGPT; GPT-4/GPT-4o; Gemini; GLM-4; DeepSeek; ERNIE-Bot; ChatGLM; local LLMs	Question answering and conceptual explanations; generation of ideas, designs, and code; feedback; conversational interaction.	17/29	58.6
Computer-vision- and machine-learning-centered recognition and execution systems	YOLOv3; YOLOR; CNN; OpenCV; Azure Custom Vision; JetBot; Jetson Nano; ESP32-CAM; Birdbuddy	Image, object, and state recognition; model training and inference; sensing and control in robotics and AIoT.	7/29	24.1
Nongenerative intelligent tutors and adaptive support systems	Transformer/RNN-based tutors; knowledge-based agents; Dialogflow chatbots; performance-prediction and adaptive systems	Diagnosis of responses and performance; hints and resources; adaptation of learning paths and task difficulty.	5/29	17.2

C. AI Use Across STEM Problem-Solving Activity Categories

For RQ1, each study could be coded in more than one activity category. Table VIII therefore reports non-mutually exclusive counts within the selected 29-study corpus. The values are descriptive of this corpus and should not be interpreted as prevalence estimates for AI use in STEM education.

TABLE VIII. REPORTED AI USE ACROSS STEM PROBLEM-SOLVING ACTIVITY CATEGORIES

Activity Category	Studies (n/29)	Percent (%)	Representative Pattern of AI Use
A1. Problem Understanding and Exploration	28/29	96.6	Conceptual explanations, examples and context, question answering, and exploration of problem requirements and relevant information.
A2. Planning and Design	19/29	65.5	Generation of solution strategies and design alternatives; review of constraints; task decomposition; and procedural planning.
A3. Implementation and Construction	22/29	75.9	Code and model implementation, device configuration, and construction using circuits, sensors, robots, or AI kits.
A4. Experimentation and Testing	20/29	69.0	Model execution, measurement and observation, examination of operation and performance, and inspection of errors and failures.
A5. Analysis and Interpretation	17/29	58.6	Analysis of data and experimental results and review and interpretation of errors, patterns, and predictions.
A6. Revision and Improvement	27/29	93.1	Error diagnosis, revision suggestions, repeated attempts, model retraining, and artifact improvement.
A7. Reflection	16/29	55.2	Self-assessment, examination of grounds for judgment, elaboration of explanations, and reflection on the problem-solving process.

Reported AI use was identified most often in Problem Understanding and Exploration (28/29 studies) and Revision and Improvement (27/29). During initial problem engagement, AI was used to explain concepts and examples and to clarify the problem context and requirements. During revision, systems identified errors or limitations in student responses, code, designs, experimental results, or constructed artifacts. Students then used feedback or revision directions to modify their work or try again.

Implementation and Construction was identified in 22 of 29 studies, and Experimentation and Testing in 20 of 29. AI-supplied information was used in constructing code, models, circuits, and devices, or it operated as part of the constructed system. In studies using computer vision, machine learning, robotics, or AIoT, model inferences were linked to sensors, motors, or controllers. Students executed these systems and examined the outputs to assess performance and failure conditions.

Planning and Design was identified in 19 of 29 studies. AI suggested solution strategies, product ideas, design alternatives, component choices, and procedural plans, or it was used to review student-defined requirements and constraints. Some systems supplied relatively complete designs; in other cases, students compared alternatives, added constraints, and revised AI-generated outputs.

Analysis and Interpretation was identified in 17 of 29 studies, and Reflection in 16 of 29. Analysis and Interpretation was coded only when students examined analytical results, predictions, errors, or patterns and constructed meaning from them; internal data processing alone was insufficient. Reflection was coded only when students used an AI-generated question or feedback for self-assessment, examination of their reasoning, or reconstruction of an explanation rather than merely reporting a general post-activity impression.

D. Functional Roles of AI

For RQ2, the analysis compared AI functions, outputs, activity contexts, and subsequent student actions across the 37 reporting-eligible cases. Four recurring roles were identified: Conceptual Resource, Generative Resource, Operational Component, and Adaptive Scaffold. Roles were non-mutually exclusive, and a case could receive more than one role when distinct output–action relationships were documented. Table IX reports counts within this analytical subset; the figures describe the reporting-eligible cases and are not estimates of the broader distribution of AI roles.

TABLE IX. RECURRING FUNCTIONAL ROLES IN THE REPORTING-ELIGIBLE CASES

Functional Role	Reporting-Eligible Cases (n/37)
F1. Conceptual Resource	11/37
F2. Generative Resource	13/37
F3. Operational Component	13/37
F4. Adaptive Scaffold	11/37

a) AI as a conceptual resource: In a Conceptual Resource case, AI supplied an explanation, example, question-and-answer exchange, contextual detail, or representational support that students used to clarify a problem, concept, operating principle, or representation. Reported cases included explanations of electronics and robotics components, connections between buoyancy and submarine design, clarification of data-visualization concepts, and images or examples for unfamiliar tools. Subsequent actions included asking follow-up questions, refining a problem representation, and applying an explanation during design or construction. The role was assigned from this use of the output, not from text generation alone.

b) AI as a generative resource: In a Generative Resource case, AI produced a candidate idea, strategy, plan, design, code fragment, model, or draft that entered the student's solution space. Reported outputs included sensor and component comparisons, project ideas, sequenced plans, circuit and structural alternatives, code drafts, and context-specific product concepts. Students subsequently compared, selected, modified, negotiated, implemented, or rejected these candidates. Some systems broadened the design space with multiple alternatives or what-if questions; others supplied a relatively complete candidate that became a reference point for later work.

c) AI as an operational component: In an Operational Component case, AI performed part of the inquiry or engineering system rather than merely supplying advice. Examples included computer vision used to assess structural stability or locate objects, image or speech classification used to trigger a device, bird classification used to accumulate ecological observations, and predictions used to control motors or LEDs. Students integrated the resulting inferences into physical or computational systems. They then observed and tested system operation and examined performance or failure conditions.

d) AI as an adaptive scaffold: In an Adaptive Scaffold case, AI analyzed a student response, code, design, experimental result, performance state, or task state and provided contingent feedback, hints, revision directions, or reflective questions. Reported cases included feedback adjusted after repeated misconceptions, artifact evaluation against criteria, code or circuit diagnosis, sensor-calibration suggestions, task adaptation, and identification of missing evidence in an argument. Students then revised an artifact, tried again, sought additional evidence, or reconsidered a judgment. This role was distinguished from general explanation by its dependence on specific evidence about the learner, artifact, or task state.

E. Relationship Between Technical Configuration and Functional Role

Functional roles did not correspond one-to-one with technology types. A role was assigned based on the relationship among the operation or output, the activity context, and the subsequent student action.

Generative-AI-based conversational and agent-based systems served as Conceptual Resources, Generative Resources, and Adaptive Scaffolds. A system served as a Conceptual Resource when it explained concepts or operating principles that students used to understand a problem or seek information. It served as a Generative Resource when it produced ideas, design alternatives, or code that students compared, selected, or modified. It served as an Adaptive Scaffold when it analyzed an existing response or artifact and identified errors, missing evidence, or directions for revision.

Computer-vision and machine-learning systems most often served as Operational Components by detecting or classifying states, objects, or patterns and linking inferences to model execution or device control. Their outputs could also serve as evidence for analysis and improvement. In cases where students observed misclassifications or low-confidence predictions and then added training images or retrained a model, the operational output became a basis for Analysis and Interpretation and Revision and Improvement.

Intelligent tutors and adaptive support systems mainly served as Conceptual Resources and Adaptive Scaffolds. They acted as Conceptual Resources when they answered learner questions or supplied relevant explanations. They acted as Adaptive Scaffolds when analysis of responses or performance histories changed the level of feedback, the subsequent task, or the learning path.

Technology type defines a range of possible functions but does not determine a functional role. Analysis of an AI-STEM environment should therefore document the AI operation and output, the activity context, and the subsequent student action at the case level rather than assign a fixed role to a product or algorithm.

V. DISCUSSION

A. Concentrations and Relative Gaps Across STEM Activities

Within the selected corpus, reported AI use was concentrated in Problem Understanding and Exploration and Revision and Improvement. This pattern may partly reflect the visibility of students' questions and information needs during initial problem engagement, which provides a clear basis for supplying concepts, examples, and contextual information. During revision, errors, incomplete artifacts, and failed results likewise provide identifiable cues for diagnosis and suggested modification [33], [38]. The pattern is consistent with reviews reporting extensive use of generative AI, intelligent tutoring, and feedback systems for explanation and error correction [1], [2], but it should be interpreted as a feature of the reviewed corpus rather than a population-level distribution.

AI also appeared widely in Implementation and Construction and Experimentation and Testing. Students tested AI operations and failure conditions, and in some cases they responded to misclassifications by adding data or retraining a model [17], [22], [34]. In these cases, AI functioned both as a tool within problem solving and as an object whose performance and limitations students investigated. This pattern is consistent with the iterative character of STEM practice, which links implementation, testing, and improvement [5], [7], [8].

Analysis and Interpretation and Reflection were reported less often. In problem-based and project-based learning, these activities help transform experience gained through performance into knowledge and strategy [9], [10]. The lower count for Reflection may reflect both system design and reporting practice. Many systems organized interaction around immediate questions, errors, and incomplete artifacts, making exploration and revision more visible than delayed metacognitive activity. The coding rule also excluded general satisfaction or post-activity impressions and required explicit evidence that students used an AI-generated prompt or feedback to examine their reasoning. Reflective activity may therefore have occurred without being captured in system logs or article reports. The finding should be read as a lower frequency of explicitly reported AI-linked reflection, not as evidence that reflection was absent. Future studies should report whether students compared evidence, justified a decision, explained why a revision was needed, and retested a result.

B. Functional Roles Beyond Technology Labels

The four recurring roles should not be read as technology categories or as a comprehensive taxonomy of AI in STEM education. They constitute a provisional analytic framework derived from 37 reporting-eligible cases. Each role describes how an AI function or output was incorporated into a particular problem-solving activity, as evidenced by the action that followed. A single system could therefore be associated with several roles across different interactions.

An LLM served as a Conceptual Resource when it explained a concept, as a Generative Resource when it proposed an idea or design, and as an Adaptive Scaffold when it analyzed a student artifact and suggested revision. Computer vision or machine learning served as an Operational Component when recognition or prediction controlled a device, whereas a resulting misclassification could also become evidence for analysis and improvement. An intelligent tutor likewise performed different roles when it supplied a general explanation and when it adjusted feedback in response to a student's performance state.

This case-level framework complements rather than replaces technology-centered reviews [3], [6]. Technology type identifies possible operations; it does not establish how an output functioned in the learner's activity. The framework's analytical value lies in separating technical capability from reported use and in requiring an explicit link to a subsequent student action. Its boundaries and reproducibility should be tested in independent corpora.

C. Connecting AI Outputs to Students' Subsequent Actions

The reporting-eligibility rule required evidence of an AI function or output, the student activity in which it was used, and a subsequent student action. This rule distinguishes a description of what AI generated or performed from a documented interaction in which the output entered students' problem-solving activity. It also means that the case-level synthesis is shaped by what the source studies reported.

An explanation, code fragment, prediction, or feedback message alone does not establish that students used it to support problem solving. A more specific account is possible when students are reported to have reviewed the output and accepted, revised, or rejected it; compared it with an alternative; or executed and tested it again. Mediation in this review therefore denotes an observable connection between an AI output and a subsequent student action, not a causal claim that AI improved learning.

Future studies should first report the state of the student or task immediately before AI use and describe the AI function and output. They should then connect those elements to the student's acceptance, modification, rejection, or retry and to any resulting change in the artifact. This sequence addresses one dimension of the methodological gap identified in previous AI-STEM research: limited interaction-level evidence [2].

Relevant evidence may include dialogue records, revision histories for code and designs, model-execution logs, experimental data, students' explanations and justifications, and subsequent performance. Such records can show whether an output entered observable activity. Establishing whether that use improved understanding, transfer, or independent problem solving requires a separate causal or longitudinal design.

D. Implications for the Design and Evaluation of AI-STEM Systems

The findings support a set of design and evaluation propositions organized around learner activity and reported functional role rather than technology type. These propositions are not effects established by the review; they specify relationships that future system evaluations can test.

First, a system may be designed to infer the current activity from recent artifacts, attempts, errors, and requests and to vary the form of support accordingly. During Analysis and Interpretation or Reflection, prompts may ask learners to compare evidence, explain their reasoning, and retest a result instead of presenting a completed answer. Empirical studies should examine whether such prompts actually elicit the intended activity.

Second, interfaces should distinguish among conceptual explanations, candidate solutions, operational outputs, and diagnoses of student work. When feasible, outputs should identify sources, assumptions, scope, and uncertainty so that learners can inspect the basis of a response.

Third, support may be staged to retain opportunities for student judgment. Learners can state an initial idea and rationale before receiving cues or feedback. Subsequent support can be adjusted or reduced only after its effects on performance and independent action have been examined. Teachers should retain authority over automated evaluation, task adaptation, and the conditions under which feedback is applied.

Fourth, evaluation should connect system outputs to subsequent student actions rather than stop at output accuracy. Linking each AI response to acceptance, modification, rejection, artifact change, or retry allows researchers to determine whether and how the output entered observable problem-solving activity. Claims about learning gains require additional outcome measures and an appropriate comparison or longitudinal design.

Accordingly, evaluation frameworks may consider technical performance, activity fit, output transparency, student judgment and revision, and evidence of increasing independence. Table X translates these findings into testable design and evaluation principles.

TABLE X. DESIGN AND EVALUATION PRINCIPLES FOR AI-STEM LEARNING SYSTEMS

Analytical Finding	Design or Evaluation Challenge	Testable Principle
Reported AI use was concentrated in Problem Understanding and Exploration and Revision and Improvement within the selected corpus.	Systems may focus on readily observable questions and errors while leaving interpretation and reflection less visible.	Include prompts intended to elicit Analysis and Interpretation and Reflection, and test whether students engage in those activities.
The same technology was associated with multiple reported functional roles.	A technology name alone does not specify the meaning of an output or the appropriate evaluation criterion.	Distinguish conceptual explanations, candidate solutions, operational outputs, and contingent diagnoses; disclose evidence, assumptions, and uncertainty.
Operational outputs were used as evidence for analysis, revision, or retry.	Technical accuracy does not establish educational value.	Assess model performance together with whether and how students interpret outputs, revise data or artifacts, and retest.
Students' subsequent actions were not consistently reported.	It is difficult to determine whether an AI output entered observable problem-solving activity.	Record each AI response together with student acceptance, modification, rejection, artifact change, and retry.
Some generative and adaptive systems supplied nearly complete responses.	Opportunities for student judgment may be reduced.	Elicit an initial student judgment, introduce support progressively, test whether it can be reduced without loss of performance, and preserve teacher oversight.

E. Limitations

This review was designed to describe where and how AI was reported within a defined journal corpus. It did not synthesize learning effects or apply a common risk-of-bias or methodological-quality appraisal across studies. The activity and role counts therefore describe the selected corpus and reporting-eligible cases; they do not establish the effectiveness, superiority, or population prevalence of any activity or role.

The search was restricted to English-language journal articles indexed in Scopus, Web of Science, or ERIC and retrieved with the specified AI and STEM terms. IEEE Xplore and the ACM Digital Library were not searched as separate platforms, and conference proceedings were excluded. Backward reference searching broadened retrieval, but technically oriented studies disseminated through computing and engineering venues may still be underrepresented.

The corpus was also temporally concentrated: 23 of the 29 studies were published in 2025 or during the partial 2026 search period. The findings should therefore be interpreted as a recent cross-section of a rapidly developing literature, not as a longitudinal account of change from 2017 to 2026.

Finally, the case-level synthesis depended on reporting detail. Only interactions in which the AI function or output, activity context, and subsequent student action were all explicit were retained, yielding 37 reporting-eligible cases. The resulting role framework may therefore reflect the reporting practices of better-documented studies and should be treated as provisional until it is tested in independent corpora. Underreporting may also have obscured reflective activity and blurred role boundaries.

VI. CONCLUSION

Within the selected 29-study corpus, AI was reported most often in Problem Understanding and Exploration and in Revision and Improvement; it was also reported in Implementation and Construction and in Experimentation and Testing. The case-level synthesis identified four recurring roles across 37 reporting-eligible interactions: Conceptual Resource, Generative Resource, Operational Component, and Adaptive Scaffold. These roles form a provisional framework for describing the reviewed cases, not a comprehensive taxonomy or estimate of role prevalence in AI-STEM education.

Whereas prior AI-STEM reviews have primarily organized research by technology type, system function, or learning outcome, the principal contribution of this review is an interaction-centered framework for examining how AI is integrated into students' problem-solving. The framework links the activity category in which AI is used, the functional role it assumes in that context, and the student judgment or action that follows. This approach shows why the same LLM, computer-vision model, or intelligent tutor may perform different roles across contexts. AI-STEM environments should therefore be described by connecting technical operation and output to activity context and subsequent student activity rather than by product names or algorithm families alone.

Future research should report the learner or task state before AI use, the function and output provided, the learner's response, and any resulting change in the artifact or line of reasoning. Dialogue records, revision histories, model logs, experimental data, and learner explanations can strengthen this chain of evidence. Experiments, longitudinal studies, discourse analysis, and performance assessments are then needed to test effects on conceptual understanding, transfer, self-regulation, and independent problem solving. Research in early childhood, primary, vocational, and diverse linguistic and cultural settings is also needed to examine the framework's scope and boundary conditions.

Designers may use the four roles to specify what kind of support is intended at a given point in the activity. Conceptual Resources can be evaluated for source transparency and support for understanding; Generative Resources for how learners compare or modify candidates; Operational Components for access to input data and failure conditions; and Adaptive Scaffolds for the evidential basis and contingency of their guidance. These are design propositions that require empirical evaluation. Teachers should retain control over the scope of automation and the conditions under which automated actions require approval.

Accordingly, evaluations of AI-STEM environments should examine how a specific AI output entered students' problem-solving activity and what students did next, rather than merely whether AI was present or whether an outcome changed. Making this relationship explicit can improve the precision of research, system design, and evaluation while preserving attention to student judgment and responsibility. Whether particular configurations preserve or strengthen that responsibility remains an empirical question.

Declaration on Generative AI

Generative AI tools were used to assist with English translation, language refinement, and manuscript formatting.

References

[1] A. I. Akhmetova, D. M. Sovetkanova, L. K. Komekbayeva, A. E. Abdrakhmanov, D. Yessenuly, and O. S. Serikova, "A systematic review of artificial intelligence in high school STEM education research," EURASIA Journal of Mathematics, Science and Technology Education, vol. 21, no. 4, art. no. em2623, 2025, doi: 10.29333/ejmste/16222.

[2] F. E. Pacheco-Olea and C. Villacís-Macías, "Artificial intelligence in STEM education: Applications, learning outcomes, and methodological gaps (2016–2025)," Frontiers in Education, vol. 11, art. no. 1791361, 2026, doi: 10.3389/feduc.2026.1791361.

[3] W. Xu and F. Ouyang, "The application of AI technologies in STEM education: A systematic review from 2011 to 2021," International Journal of STEM Education, vol. 9, art. no. 59, 2022, doi: 10.1186/s40594-022-00377-5.

[4] T. R. Kelley and J. G. Knowles, "A conceptual framework for integrated STEM education," International Journal of STEM Education, vol. 3, art. no. 11, 2016, doi: 10.1186/s40594-016-0046-z.

[5] National Research Council, A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC, USA: National Academies Press, 2012, doi: 10.17226/13165.

[6] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—Where are the educators?" International Journal of Educational Technology in Higher Education, vol. 16, art. no. 39, 2019, doi: 10.1186/s41239-019-0171-0.

[7] NGSS Lead States, Next Generation Science Standards: For States, By States. Washington, DC, USA: National Academies Press, 2013.

[8] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, "Engineering design thinking, teaching, and learning," Journal of Engineering Education, vol. 94, no. 1, pp. 103–120, 2005, doi: 10.1002/j.2168-9830.2005.tb00832.x.

[9] C. E. Hmelo-Silver, "Problem-based learning: What and how do students learn?" Educational Psychology Review, vol. 16, no. 3, pp. 235–266, 2004, doi: 10.1023/B:EDPR.0000034022.16470.f3.

[10] J. S. Krajcik and P. C. Blumenfeld, "Project-based learning," in The Cambridge Handbook of the Learning Sciences, R. K. Sawyer, Ed. New York, NY, USA: Cambridge University Press, 2006, pp. 317–334, doi: 10.1017/CBO9780511816833.020.

[11] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, art. no. n71, 2021, doi: 10.1136/bmj.n71.

[12] J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104.

[13] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, U.K., 2014, pp. 1–10, doi: 10.1145/2601248.2601268.

[14] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," Qualitative Health Research, vol. 15, no. 9, pp. 1277–1288, 2005, doi: 10.1177/1049732305276687.

[15] N. Yannier, S. E. Hudson, and K. R. Koedinger, "Active learning is about more than hands-on: A mixed-reality AI system to support STEM education," International Journal of Artificial Intelligence in Education, vol. 30, pp. 74–96, 2020, doi: 10.1007/s40593-020-00194-3.

[16] S.-Y. Chen, W.-C. Chen, and C.-F. Lai, "Generative AI as a reflective scaffold in a UAV-based STEM project: A mixed-methods study on students' higher-order thinking and cognitive transformation," Education and Information Technologies, vol. 30, pp. 24787–24814, 2025, doi: 10.1007/s10639-025-13758-4.

[17] J. Liao, J. Yang, and W. Zhang, "The student-centered STEM learning model based on artificial intelligence project: A case study on intelligent car," International Journal of Emerging Technologies in Learning, vol. 16, no. 21, pp. 100–120, 2021, doi: 10.3991/ijet.v16i21.25105.

[18] L. Gao, M. S.-Y. Jong, C. S. Chai, and K. Li, "STEM education: Understanding secondary students' epistemic cognition in the design process with the support of a personalized multi-agent system," Computers & Education, vol. 242, art. no. 105504, 2026, doi: 10.1016/j.compedu.2025.105504.

[19] M. Wang, J. Zhu, G.-J. Hwang, S.-C. Chang, Q.-F. Yang, and D. Zhang, "Boosting student engagement in STEM: Integrating large language model-based virtual agents into alternate reality games," Journal of Computer Assisted Learning, vol. 41, art. no. e70139, 2025, doi: 10.1111/jcal.70139.

[20] V. Štuikys, R. Burbaitė, M. Binkis, and G. Ziberkas, "Developing problem-solving skills to support sustainability in STEM education using generative AI tools," Sustainability, vol. 17, no. 15, art. no. 6935, 2025, doi: 10.3390/su17156935.

[21] C. Lee and H.-J. So, "Learners leveraging generative AI for creative problem solving: Focusing on the PISA 2022 creative thinking problems," Educational Technology & Society, vol. 28, no. 4, pp. 241–258, 2025, doi: 10.30191/ETS.202510_28(4).SP04.

[22] C.-H. Lin, C.-C. Yu, P.-K. Shih, and L. Y. Wu, "STEM-based artificial intelligence learning in general education for non-engineering undergraduate students," Educational Technology & Society, vol. 24, no. 3, pp. 224–237, 2021.

[23] C.-J. Lin, H.-Y. Lee, W.-S. Wang, Y.-M. Huang, and T.-T. Wu, "Enhancing reflective thinking in STEM education through experiential learning: The role of generative AI as a learning aid," Education and

Information Technologies, vol. 30, pp. 6315–6337, 2025, doi: 10.1007/s10639-024-13072-5.

[24] Q. Guo, J. Zhen, F. Wu, Y. He, and C. Qiao, “Can students make STEM progress with large language models (LLMs)? An empirical study of LLMs integration within middle school science and engineering practice,” Journal of Educational Computing Research, vol. 63, no. 2, pp. 372–405, 2025, doi: 10.1177/07356331241312365.

[25] M. Serik, K. Tleuzhanova, and S. Nurgaliyeva, “Enhancing master’s-level STEM education through AI-driven IoT projects: A Kazakhstan experiment,” International Journal of Information and Education Technology, vol. 15, no. 10, pp. 2086–2094, 2025, doi: 10.18178/ijiet.2025.15.10.2407.

[26] L. Li, “From tool to architect: Constructing a generative AI-empowered ‘Perception–Diagnosis–Critical Thinking’ triadic teaching model in inorganic chemistry,” Journal of Science Education and Technology, early access, 2026, doi: 10.1007/s10956-026-10328-2.

[27] H.-C. Hung and P.-J. Lai, “Enhancing graduate students’ data visualization literacy through an inquiry-based learning companion system integrated with generative AI,” Educational Technology & Society, vol. 28, no. 4, pp. 304–320, 2025, doi: 10.30191/ETS.202510_28(4).SP07.

[28] R. Gabrovšek and D. Rihtaršič, “Custom generative artificial intelligence tutors in action: An experimental evaluation of prompt strategies in STEM education,” Sustainability, vol. 17, no. 21, art. no. 9508, 2025, doi: 10.3390/su17219508.

[29] C.-J. Lin, W.-S. Wang, H.-Y. Lee, P.-H. Li, Y.-M. Huang, and T.-T. Wu, “Advancing self-directed learning in STEM education: Integrating GPT-based learning aid with multimodal learning analytics,” Journal of Research on Technology in Education, early access, May 19, 2025, doi: 10.1080/15391523.2025.2504358.

[30] L. S. Ferro, F. Sapio, A. Terracina, M. Temperini, and M. Mecella, “Gea2: A serious game for technology-enhanced learning in STEM,” IEEE Transactions on Learning Technologies, vol. 14, no. 6, pp. 723–739, 2021, doi: 10.1109/TLT.2022.3143519.

[31] T.-T. Wu, H.-Y. Lee, P.-H. Chen, C.-J. Lin, and Y.-M. Huang, “Integrating peer assessment cycle into ChatGPT for STEM education: A randomised controlled trial on knowledge, skills, and attitudes enhancement,” Journal of Computer Assisted Learning, vol. 41, art. no. e13085, 2025, doi: 10.1111/jcal.13085.

[32] W. Villegas-Ch, D. Buenano-Fernandez, A. Maldonado Navarro, and A. Mera-Navarrete, “Adaptive intelligent tutoring systems for STEM education: Analysis of the learning impact and effectiveness of personalized feedback,” Smart Learning Environments, vol. 12, art. no. 41, 2025, doi: 10.1186/s40561-025-00389-y.

[33] C.-J. Sui, C.-Y. Chang, and M.-H. Yen, “STEM-5E socio-scientific argumentation with generative AI-driven scaffolding: Exploring the interplay between epistemic beliefs and learning outcomes,” Journal of Science Education and Technology, vol. 35, no. 2, pp. 429–447, 2026, doi: 10.1007/s10956-025-10257-6.

[34] M. Lohakan and C. Seetao, “Large-scale experiment in STEM education for high school students using artificial intelligence kit based on computer vision and Python,” Heliyon, vol. 10, art. no. e31366, 2024, doi: 10.1016/j.heliyon.2024.e31366.

[35] C. B. Omeh and M. A. Ayanwale, “Artificial intelligence meets PBL: Transforming computer-robotics programming motivation and engagement,” Frontiers in Education, vol. 10, art. no. 1674320, 2025, doi: 10.3389/feduc.2025.1674320.

[36] M. Tripepi, J. Yang, and D. Tariq, “Birdbuddy in the classroom: Leveraging AI-powered bird feeders for undergraduate biology education,” Journal of Microbiology & Biology Education, vol. 27, no. 1, art. no. e00223-25, 2026, doi: 10.1128/jmbe.00223-25.

[37] J. Liang and W. Yang, “Using generative artificial intelligence to facilitate early STEM learning: A case study in a Chinese kindergarten,” Early Education and Development, early access, May 6, 2026, doi: 10.1080/10409289.2026.2669922.

[38] T. El Fathi, A. Saad, H. Larhzil, D. Lamri, and E. M. Al Ibrahmi, “Integrating generative AI into STEM education: Enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance,” Disciplinary and Interdisciplinary Science Education Research, vol. 7, art. no. 6, 2025, doi: 10.1186/s43031-025-00125-z.

[39] A. A. Zaus, M. Giatman, M. A. Zaus, A. A. Zaus, and S. Islami, “Enhancing critical thinking and teamwork through AI-integrated project-based learning in engineering education,” International Journal on Informatics Visualization, vol. 10, no. 1, pp. 94–104, Jan. 2026, doi: 10.62527/joiv.10.1.3515.

[40] W.-S. Wang, C.-J. Lin, H.-Y. Lee, Y.-M. Huang, and T.-T. Wu, “Integrating feedback mechanisms and ChatGPT for VR-based experiential learning: Impacts on reflective thinking and AIoT physical hands-on tasks,” Interactive Learning Environments, vol. 33, no. 2, pp. 1770–1787, 2025, doi: 10.1080/10494820.2024.2375644.

[41] T.-T. Wu, H.-Y. Lee, W.-S. Wang, C.-J. Lin, and Y.-M. Huang, “Leveraging computer vision for adaptive learning in STEM education: Effect of engagement and self-efficacy,” International Journal of Educational Technology in Higher Education, vol. 20, art. no. 53, 2023, doi: 10.1186/s41239-023-00422-5.

[42] J.-W. Fang, X.-G. Guo, X.-P. Meng, G.-J. Hwang, and Y.-F. Tu, “Realization of parasocial interaction theory in STEAM education with an AI pedagogical agent: Insights from learning performance, collaboration tendency, and problem-solving tendency,” Interactive Learning Environments, vol. 34, no. 2, pp. 694–717, 2026, doi: 10.1080/10494820.2025.2508326.

[43] Y. Ji, Z. Zhan, T. Li, X. Zou, and S. Lyu, “Human–machine cocreation: The effects of ChatGPT on students’ learning performance, AI awareness, critical thinking, and cognitive load in a STEM course toward entrepreneurship,” IEEE Transactions on Learning Technologies, vol. 18, pp. 402–415, 2025, doi: 10.1109/TLT.2025.3554584.