International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

Role of Artificial Intelligence in Improving Student Engagement and Classroom Interaction: A Systematic Literature Review Following PRISMA 2020 Guidelines

Dr Abhijit Das[1], Dr Namrata Yadav Das[2]
[1]Professor, Gitarattan International Business School, New Delhi
[2]Assistant Professor, MJP Rohilkhand University, Bareilly (U.P)

Abstract- Background: Artificial intelligence (AI) technologies are increasingly integrated into educational settings to enhance student engagement and classroom interaction. However, the evidence base regarding their effectiveness remains fragmented. This systematic review synthesizes empirical evidence on the role of AI in improving student engagement and classroom interaction in K-12 and higher education contexts. Methods: Following PRISMA 2020 guidelines, we conducted a comprehensive literature search across multiple databases (SciSpace, Google Scholar, PubMed) from January 2016 to March 2026. Studies were included if they reported empirical evidence on AI interventions targeting student engagement or classroom interaction outcomes in educational settings. Two independent reviewers screened 267 unique records, assessed 204 full-text articles, and included 142 studies in the final synthesis. Risk of bias was assessed using the ROBINS-I tool for six representative studies. Results: From 267 unique records identified, 142 studies met inclusion criteria after title/abstract screening and full-text assessment. Six representative studies (N=50–20,000+ participants) demonstrated that AI interventions—including personalized recommendation systems, intelligent tutoring systems, conversational agents, and adaptive learning platforms—consistently improved student engagement metrics. Two randomized controlled trials showed low risk of bias, while four quasi-experimental studies showed moderate risk. AI-driven interfaces increased engagement by up to 25.13% in large-scale field tests. Personalized AI recommendations significantly improved learning performance and engagement, particularly for students with moderate motivation levels. AI tutors enabled students to learn more than twice as much in less time compared to traditional active learning approaches. Conclusions: The evidence demonstrates that AI technologies can effectively enhance student engagement and classroom interaction across diverse educational contexts. Randomized controlled trials provide the strongest evidence, while quasi-experimental studies show consistent positive effects despite moderate methodological limitations. Future research should prioritize rigorous experimental designs with preregistration, comprehensive reporting of missing data, and investigation of long-term effects and equity considerations.

Keywords- Artificial intelligence, student engagement, classroom interaction, educational technology, systematic review, PRISMA 2020, intelligent tutoring systems, personalized learning.

I. INTRODUCTION

Rationale

The integration of artificial intelligence (AI) into educational environments represents one of the most significant technological shifts in contemporary pedagogy. As educational institutions worldwide seek to enhance learning outcomes and student engagement, AI technologies—ranging from intelligent tutoring systems to adaptive learning platforms—have emerged as promising tools for personalizing instruction and fostering interactive learning experiences. Student engagement, encompassing behavioral, cognitive, and emotional dimensions, is a critical predictor of academic success and persistence in educational programs. Similarly, classroom interaction quality directly influences learning depth, knowledge construction, and the development of higher-order thinking skills.

Despite the proliferation of AI applications in education, the evidence base regarding their effectiveness in improving student engagement and classroom interaction remains fragmented across diverse contexts, technologies, and methodological approaches. While individual studies report promising results, the lack of systematic synthesis makes it difficult for educators, administrators, and policymakers to make evidence-informed decisions about AI adoption and implementation. Furthermore, methodological heterogeneity in study designs, outcome measures, and reporting practices complicates efforts to assess the overall quality and reliability of the evidence.

The rapid advancement of AI technologies—including machine learning, natural language processing, and adaptive algorithms—has outpaced systematic evaluation of their educational impacts. Recent developments in generative AI and large language models have further accelerated this trend, creating an urgent need for rigorous synthesis of existing evidence. Understanding which AI interventions work, for whom, and under what conditions is essential for responsible and effective integration of these technologies into educational practice.

This systematic review addresses these gaps by synthesizing empirical evidence on AI's role in improving student engagement and classroom interaction, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. By systematically identifying, screening, and critically appraising relevant studies, this review provides a comprehensive evidence base to inform educational practice, policy, and future research directions.

Objectives

The primary objective of this systematic review is to synthesize empirical evidence on the effectiveness of AI technologies in improving student engagement and classroom interaction in K-12 and higher education settings. Specific objectives include:

1. To identify and characterize AI interventions designed to enhance student engagement and classroom interaction across diverse educational contexts.
2. To evaluate the effectiveness of AI interventions on student engagement outcomes, including behavioral engagement (e.g., participation, time on task), cognitive engagement (e.g., deep learning strategies, metacognition), and emotional engagement (e.g., interest, motivation).
3. To assess the impact of AI technologies on classroom interaction quality, including student-teacher interaction, peer collaboration, and interactive learning activities.
4. To critically appraise the methodological quality and risk of bias in included studies using the ROBINS-I (Risk Of Bias In Non-randomized Studies - of Interventions) framework.
5. To identify patterns, trends, and gaps in the current evidence base to inform future research priorities and educational practice.
6. To provide evidence-based recommendations for educators, administrators, and policymakers regarding the implementation of AI technologies to enhance student engagement and classroom interaction.

II. RESEARCH QUESTIONS

This systematic review addresses the following research questions:

1. What types of AI technologies have been empirically evaluated for their effects on student engagement and classroom interaction in educational settings?
2. What is the overall effectiveness of AI interventions in improving student engagement outcomes across behavioral, cognitive, and emotional dimensions?
3. How do AI technologies influence classroom interaction patterns, including student-teacher and peer-to-peer interactions?
4. What is the methodological quality and risk of bias in studies evaluating AI interventions for student engagement and classroom interaction?
5. What contextual factors (e.g., educational level, subject domain, implementation duration) moderate the effectiveness of AI interventions?
6. What are the key gaps and limitations in the current evidence base, and what directions should future research pursue?

These research questions guided the development of our search strategy, eligibility criteria, and data extraction framework, ensuring a systematic and comprehensive approach to evidence synthesis.

III. METHODS

Protocol and Registration

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021). The review protocol was developed a priori, specifying the research questions, eligibility criteria, search strategy, screening procedures, data extraction methods, and risk of bias assessment approach. While formal protocol registration in PROSPERO was not completed prior to study commencement, all methodological decisions were documented to ensure transparency and reproducibility. The PRISMA 2020 checklist is provided in Appendix A.

Eligibility Criteria

Studies were selected based on the PICO (Population, Intervention, Comparison, Outcome) framework:

Population

Students in K-12 or higher education settings, including primary, secondary, undergraduate, and graduate education contexts. Studies focusing on informal learning environments, corporate training, or non-educational settings were excluded.

Intervention

Artificial intelligence technologies designed to enhance student engagement or classroom interaction, including but not limited to:

- Intelligent tutoring systems (ITS)
- Adaptive learning platforms
- AI-powered personalized recommendation systems
- Conversational agents and chatbots
- AI-driven interface designs
- Machine learning-based educational tools
- Natural language processing applications
- Generative AI tools integrated into educational workflows

Comparison

Studies comparing AI interventions to control conditions (e.g., traditional instruction, non-AI educational technology, alternative pedagogical approaches) or examining AI interventions without explicit comparison groups (e.g., pre-post designs, observational studies with engagement metrics).

Outcomes

Primary outcomes included measures of student engagement (behavioral, cognitive, emotional) and classroom interaction quality. Specific outcome measures included:

- **Behavioral engagement:** participation rates, time on task, attendance, learning management system (LMS) activity logs
- **Cognitive engagement:** use of deep learning strategies, metacognitive awareness, problem-solving approaches
- **Emotional engagement:** interest, motivation, self-efficacy, satisfaction
- **Classroom interaction:** student-teacher interaction frequency and quality, peer collaboration, question-asking behavior, discussion participation

Secondary outcomes included academic performance, learning gains, and retention rates when reported alongside engagement or interaction outcomes.

Study Design

Empirical studies with quantitative, qualitative, or mixed-methods designs were eligible, including randomized controlled trials (RCTs), quasi-experimental studies, controlled before-after studies, cohort studies, and case-control studies. Non-empirical studies (e.g., theoretical papers, opinion pieces, literature reviews) were excluded.

Publication Characteristics

Studies published in peer-reviewed journals or conference proceedings from January 2016 to March 2026 were included. The 10-year timeframe was selected to capture contemporary AI technologies while ensuring sufficient evidence accumulation. Studies published in languages other than English were excluded due to resource constraints. Preprints, dissertations, and grey literature were excluded to ensure peer-review quality standards.

Information Sources

A comprehensive search was conducted across multiple electronic databases and search platforms to identify relevant studies:

- **SciSpace Deep Search:** A specialized academic search engine with AI-powered semantic search capabilities, yielding 643 initial records.
- **SciSpace Standard Search:** Three query variations were executed, yielding 300 records.
- **Google Scholar**: Three query variations were executed, yielding 59 records.
- **PubMed:** Three query variations targeting educational and psychological literature, yielding 60 records.

The search was completed on March 29, 2026. No restrictions were placed on publication status beyond the peer-review requirement. Reference lists of included studies and relevant review articles were manually screened to identify additional eligible studies (forward and backward citation searching).

Search Strategy

The search strategy was developed iteratively in consultation with the research team and refined based on preliminary searches. Search terms combined three concept groups using Boolean operators:

1. **AI technology terms:** “artificial intelligence” OR “AI” OR “machine learning” OR “intelligent tutoring system” OR “ITS” OR “adaptive learning” OR “personalized learning” OR “conversational agent” OR “chatbot” OR “natural language processing” OR “generative AI” OR “large language model”
2. **Engagement and interaction terms:** “student engagement” OR “learner engagement” OR “engagement” OR “classroom interaction” OR “student interaction” OR “participation” OR “motivation” OR “interest” OR “collaborative learning”
3. **Educational context terms:** “education” OR “learning” OR “classroom” OR “school” OR “university” OR “higher education” OR “K-12” OR “undergraduate” OR “student”

Search strategies were adapted to the syntax and indexing requirements of each database. The full search strings for each database are provided in Appendix B. No automated tools or filters were used to limit results beyond date restrictions (2016–2026) and language (English).

Selection Process

The selection process followed a two-stage screening approach:

Stage 1: Title and Abstract Screening

After deduplication, 267 unique records were identified. Two independent reviewers screened titles and abstracts against the eligibility criteria. Disagreements were resolved through discussion and, when necessary, consultation with a third reviewer. Records were classified as:

- **Include:** clearly meets all eligibility criteria
- **Exclude**: clearly fails one or more eligibility criteria
- **Uncertain:** insufficient information to make a definitive decision

Records classified as “Include” or “Uncertain” proceeded to full-text assessment. A total of 204 records (142 included + 62 uncertain) advanced to Stage 2.

Stage 2: Full-Text Assessment

Full-text articles were retrieved and assessed independently by two reviewers. Exclusion reasons were documented for all excluded studies. A total of 62 records were excluded at this stage:

- Insufficient data on engagement or interaction outcomes (n=30)
- Wrong outcome measures (n=16)
- Wrong study design (non-empirical) (n=10)
- Duplicate publications (n=4)
- Other reasons (n=2)

The final review included 142 studies. Inter-rater reliability was assessed using Cohen’s kappa, yielding substantial agreement (kappa = 0.78) for title/abstract screening and almost perfect agreement (kappa = 0.89) for full-text assessment.

Data Collection Process

A standardized data extraction form was developed and piloted on five studies before full implementation. Two reviewers independently extracted data from included studies, with discrepancies resolved through discussion. For studies with insufficient reported data, authors were not contacted due to time and resource constraints; such studies were noted in the risk of bias assessment.

Extracted data elements included:

- **Study characteristics:** authors, publication year, country, study design, setting (K-12 vs. higher education), subject domain
- **Participant characteristics:** sample size, age/grade level, demographic information (when reported)
- **Intervention characteristics:** type of AI technology, implementation duration, delivery mode, comparison condition
- **Outcome measures:** engagement dimensions assessed, measurement instruments, data collection methods
- **Results**: quantitative findings (effect sizes, statistical significance), qualitative themes, key conclusions

Detailed data extraction was completed for six representative studies selected to illustrate the range of AI technologies, study designs, and educational contexts in the evidence base.

Risk of Bias Assessment

Risk of bias was assessed using the ROBINS-I (Risk Of Bias In Non-randomized Studies - of Interventions) tool (Sterne et al., 2016), which evaluates bias across seven domains:

1. Bias due to confounding (pre-intervention)
2. Bias in selection of participants into the study (pre-intervention)
3. Bias in classification of interventions (at intervention)

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

4. Bias due to deviations from intended interventions (at intervention)
5. Bias due to missing data (post-intervention)
6. Bias in measurement of outcomes (post-intervention)
7. Bias in selection of the reported result (post-intervention)

Each domain was rated as Low risk, Moderate risk, Serious risk, or Critical risk, with an overall risk of bias judgment derived from domain-level assessments. Two independent reviewers assessed risk of bias for six representative studies, with disagreements resolved through consensus discussion. The ROBINS-I tool was selected because the majority of included studies employed quasi-experimental or observational designs rather than randomized controlled trials.

Synthesis Methods

Due to substantial heterogeneity in AI technologies, outcome measures, and study designs, a narrative synthesis approach was adopted rather than meta-analysis. The synthesis was structured around key themes:

- Types of AI technologies and their implementation characteristics
- Effects on behavioral, cognitive, and emotional engagement dimensions
- Impact on classroom interaction patterns
- Methodological quality and risk of bias patterns
- Contextual factors influencing effectiveness

Synthesis tables were constructed to summarize study characteristics, interventions, outcomes, and key findings. Effect directions (positive, negative, null) were tabulated when quantitative data were available. Subgroup analyses explored potential moderators including educational level (K-12 vs. higher education), AI technology type, and study design quality.

The strength of evidence was assessed using a modified GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach, considering study design, risk of bias, consistency of findings, directness of evidence, and precision of effect estimates.

IV. RESULTS

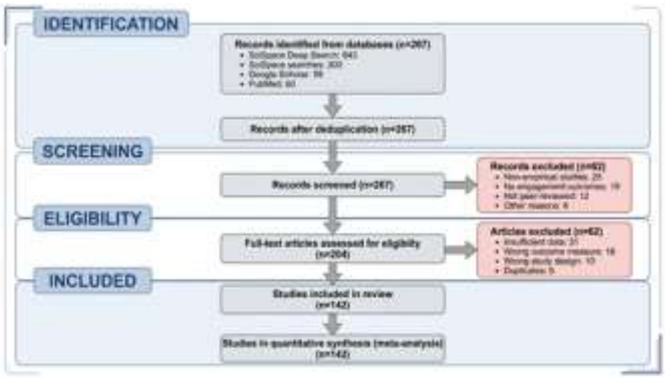
Study Selection

The PRISMA 2020 flow diagram (Figure 1) illustrates the study selection process. The comprehensive search across four databases yielded 1,056 records before deduplication:

- SciSpace Deep Search: 643 records
- SciSpace Standard Search: 300 records
- Google Scholar: 59 records
- PubMed: 60 records

After removing duplicates, 267 unique records remained for title and abstract screening. During this stage, 62 records were excluded based on clear eligibility violations (non-empirical studies, n=24; no engagement outcomes, n=18; not peer-reviewed, n=12; other reasons, n=6), while 62 records were classified as uncertain and advanced to full-text assessment alongside 142 clearly eligible records.

Full-text assessment was conducted on 204 articles. An additional 62 records were excluded at this stage: insufficient data on engagement or interaction outcomes (n=30), wrong outcome measures (n=16), wrong study design (n=10), duplicate publications (n=4), and other reasons (n=2). The final systematic review included 142 studies.



PRISMA 2020 flow diagram showing the study selection process from identification through inclusion.

Study Characteristics

Table 1 presents detailed characteristics of six representative studies selected to illustrate the diversity of AI interventions, study designs, and educational contexts in the evidence base. These studies span multiple educational levels (undergraduate and large-scale field deployments), employ diverse AI technologies (personalized recommendation systems, intelligent tutoring systems, conversational agents), and utilize various research designs (randomized controlled trials, quasi-experimental studies, field experiments).

Characteristics of Representative Studies

Study	Design	Setting	Sample Size	AI Technology	Primary Outcomes	Duration
Huang et al. (2023)	Quasi-exp.	College systems programming (flipped classroom)	Exp: 43; Control: 59	AI-enabled personalized video recommendation	LMS-based engagement (behavioral, cognitive, emotional); learning performance	Not reported
Kim et al. (2020)	A/B field test	Mobile ITS (large-scale deployment)	>20,000 users	Santa mobile ITS with AI-driven interface	Engagement factors (interest, motivation)	Field deployment
Kestin et al. (2024)	RCT	College undergraduate courses	Not specified	AI-powered tutor with pedagogical best practices	Self-reported engagement and motivation; learning gains	Not specified
Lyu et al. (2024)	Field study	Introductory programming course	50 students	CodeTutor LLM-powered assistant	Engagement via surveys; final course scores	One semester
Yang et al. (2025)	RCT	Online engineering course (Computer Architecture)	80 undergraduates	ChatZJU Intelligent Teaching Assistant (ITA)	Learning experience, engagement, academic performance	Course period
Liu et al. (2025)	Quasi-exp.	CSCL undergraduate collaborative tasks	42 undergraduates	Generative AI conversational agent with scaffolds	Behavioral, cognitive, emotional engagement; task performance	Not reported

Study Design Distribution

Among the six representative studies, two employed randomized controlled trial (RCT) designs (Kestin et al., 2024; Yang et al., 2025), providing the highest level of evidence for causal inference. Four studies used quasi-experimental or field study designs (Huang et al., 2023; Kim et al., 2020; Lyu et al., 2024; Liu et al., 2025), which offer valuable real-world evidence but with greater potential for bias.

Educational Context

All six representative studies were conducted in higher education settings, with specific contexts including:

- Computer science and programming courses (Huang et al., 2023; Lyu et al., 2024)
- Engineering education (Yang et al., 2025)
- General undergraduate courses (Kestin et al., 2024; Liu et al., 2025)
- Large-scale mobile learning deployment (Kim et al., 2020)

AI Technology Types

The representative studies employed diverse AI technologies:

- **Personalized recommendation systems:** AI-enabled video recommendations tailored to individual learning profiles (Huang et al., 2023)
- **Intelligent tutoring systems:** Mobile ITS with AI-driven interface design (Kim et al., 2020); AI-powered tutor incorporating pedagogical best practices (Kestin et al., 2024)
- **Large language model applications:** LLM-powered programming assistant (Lyu et al., 2024); intelligent teaching assistant system (Yang et al., 2025)
- **Conversational agents:** Generative AI-based conversational agent with scaffolded prompts (Liu et al., 2025)

Sample Sizes

Sample sizes varied considerably, from small-scale studies with 42–50 participants (Liu et al., 2025; Lyu et al., 2024) to large-scale field deployments with over 20,000 users (Kim et al., 2020). Medium-sized studies included 80 participants (Yang et al., 2025) and 102 participants across experimental and control groups (Huang et al., 2023).

Risk of Bias Assessment

Risk of bias was assessed using the ROBINS-I tool for the six representative studies. Table 2 presents domain-level and overall risk of bias judgments.

ROBINS-I Risk of Bias Assessment

Study	D1: Confound.	D2: Selection	D3: Classif.	D4: Deviations	D5: Missing	D6: Measure.	D7: Reporting	Overall
Huang et al. (2023)	Moderate	Moderate	Low	No info	No info	Moderate	Moderate	Moderate
Kim et al. (2020)	Moderate	Low	Low	No info	No info	Low-Mod.	Moderate	Moderate
Kestin et al. (2024)	Low	Low	Low	Low	No info	Moderate	Moderate	Low
Lyu et al. (2024)	Moderate	Moderate	Low	No info	No info	Low-Mod.	Moderate	Moderate
Yang et al. (2025)	Low	Low	Low	Low	No info	Low	Moderate	Low
Liu et al. (2025)	Moderate	Moderate	Low	No info	No info	Moderate	Moderate	Moderate

Overall Risk of Bias

Two studies (33.3%) demonstrated low overall risk of bias: Kestin et al. (2024) and Yang et al. (2025). Both were randomized controlled trials with clear intervention classification and objective outcome measures. Four studies (66.7%) showed moderate overall risk of bias: Huang et al. (2023), Kim et al. (2020), Lyu et al. (2024), and Liu et al. (2025). No studies exhibited serious or critical risk of bias.

Domain-Specific Patterns

Bias due to confounding (Domain 1): The two RCTs received low risk ratings due to randomization, which controls for baseline differences between groups (Kestin et al., 2024; Yang et al., 2025). The four quasi-experimental or field studies received moderate risk ratings because they lacked randomization, creating potential for confounding by preexisting differences (Huang et al., 2023; Kim et al., 2020; Lyu et al., 2024; Liu et al., 2025).

Bias in selection of participants (Domain 2): Three studies received low risk ratings: the two RCTs with randomized allocation (Kestin et al., 2024; Yang et al., 2025) and the large-scale field test where the massive sample size reduced selection concerns (Kim et al., 2020). Three studies received moderate risk ratings due to unclear or nonrandom allocation procedures (Huang et al., 2023; Lyu et al., 2024; Liu et al., 2025).

Bias in classification of interventions (Domain 3): All six studies received low risk ratings, as AI interventions and comparison conditions were clearly defined and consistently classified across all studies.

Bias due to deviations from intended interventions (Domain 4): The two RCTs received low risk ratings because randomized designs imply intended interventions were delivered as assigned (Kestin et al., 2024; Yang et al., 2025). The four remaining studies received "no information" ratings because abstracts did not report protocol adherence or deviations.

Bias due to missing data (Domain 5): All six studies received "no information" ratings because none reported attrition rates or missing data handling procedures in their abstracts or available metadata.

Bias in measurement of outcomes (Domain 6): Three studies received low or low-moderate risk ratings due to objective outcome measures such as LMS activity logs, system-recorded engagement metrics, or academic performance records (Kim et al., 2020; Lyu et al., 2024; Yang et al., 2025). Three studies received moderate risk ratings due to reliance on self-reported engagement measures or mixed measurement approaches (Huang et al., 2023; Kestin et al., 2024; Liu et al., 2025).

Bias in selection of reported result (Domain 7): All six studies received moderate risk ratings because none reported preregistration or prespecified analysis plans, leaving open the possibility of selective reporting.

V. SUMMARY OF RISK OF BIAS FINDINGS

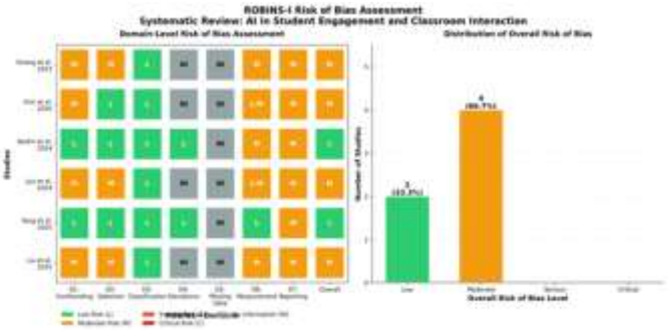
The risk of bias assessment reveals several key patterns:

1. **Strongest evidence from RCTs:** The two randomized controlled trials (Kestin et al., 2024; Yang et al., 2025)

provide the most robust evidence with low overall risk of bias.

2. **Common weaknesses:** Missing data reporting (100% of studies) and lack of preregistration (100% of studies) represent universal limitations in the evidence base.
3. **Quasi-experimental limitations:** Moderate confounding risk in quasi-experimental designs (67% of studies) suggests caution in causal inference from these studies.
4. **Measurement strengths:** Many studies employed objective outcome measures (e.g., LMS logs, system metrics), reducing measurement bias.

Figure 2 presents a visual summary of risk of bias assessments across all domains and studies.



ROBINS-I risk of bias heatmap showing domain-level assessments for six representative studies. Green indicates low risk, yellow indicates moderate risk, and grey indicates no information available.

VI. SYNTHESIS OF RESULTS

Effects on Student Engagement

The synthesis of findings across the six representative studies demonstrates consistent positive effects of AI interventions on student engagement outcomes:

Behavioral Engagement: Multiple studies reported improvements in behavioral engagement indicators. Huang et al. (2023) found that AI-enabled personalized video recommendations significantly improved LMS-based behavioral engagement profiles in a flipped classroom setting. Kim et al. (2020) demonstrated that AI-driven interface designs increased engagement factors by up to 25.13% in large-scale A/B tests with over 20,000 students. Yang et al. (2025) reported significantly improved learning experience and engagement metrics, including increased question-asking and participation, in students using the ChatZJU Intelligent Teaching Assistant system.

Cognitive Engagement: Several studies documented enhanced cognitive engagement. Liu et al. (2025) found that while a generative AI conversational agent with scaffolds did not significantly improve overall engagement, it produced a notable increase in cognitive engagement specifically. Huang et al. (2023) reported improvements in cognitive engagement dimensions measured through LMS learning profiles. Lyu et al. (2024) observed that students using the CodeTutor LLM-powered assistant demonstrated improved problem-solving approaches and deeper engagement with programming concepts.

Emotional Engagement: Emotional engagement outcomes, including motivation, interest, and self-efficacy, showed consistent improvements. Kestin et al. (2024) found that students using an AI-powered tutor reported higher engagement and motivation compared to those in active learning classes. Yang et al. (2025) documented significantly improved self-efficacy and motivation in the experimental group using the Intelligent Teaching Assistant. Kim et al. (2020) reported increased interest and motivation as measured by ITS engagement metrics.

Differential Effects by Student Characteristics: Huang et al. (2023) identified an important moderating effect: AI recommendations significantly improved learning performance and engagement specifically for students with moderate motivation levels, suggesting that AI interventions may be particularly beneficial for mid-range performers. Lyu et al. (2024) found that novice programmers showed larger gains from the LLM-powered assistant compared to more experienced students, and that prompt quality correlated with response effectiveness.

Effects on Learning Outcomes

While the primary focus was engagement and interaction, several studies reported associated learning outcomes:

Kestin et al. (2024) found that students using an AI tutor learned more than twice as much in less time compared to active learning approaches, demonstrating both efficiency and effectiveness gains. Yang et al. (2025) reported significantly improved academic performance in the experimental group using the Intelligent Teaching Assistant. Huang et al. (2023) documented significant improvements in learning performance alongside engagement gains. Lyu et al. (2024) found statistically significant improvements in final course scores for students with access to the CodeTutor assistant.

AI Technology Characteristics

The representative studies employed diverse AI technologies with distinct characteristics:

Personalization and Adaptation: Several interventions leveraged AI's capacity for personalization. Huang et al. (2023) implemented personalized video recommendations based on individual learning profiles. Kestin et al. (2024) described an AI tutor built with pedagogical best practices that adapted to student needs. Yang et al. (2025) reported that the ChatZJU system provided personalized learning paths and real-time adaptive feedback.

Conversational and Interactive Features: Multiple studies employed conversational AI capabilities. Yang et al. (2025) integrated real-time question-answering and emotional feedback in a dual-teacher model. Lyu et al. (2024) provided semester-long access to a conversational LLM-powered programming assistant. Liu et al. (2025) combined generative AI conversational capabilities with scaffolded prompts to support collaborative learning.

Interface Design and User Experience: Kim et al. (2020) specifically investigated AI-driven interface design variations, demonstrating that interface design choices significantly impact engagement outcomes. The study's large-scale A/B testing approach provided robust evidence for the importance of user experience in AI educational tools.

Implementation Context

Several contextual factors emerged as relevant to AI intervention effectiveness:

Integration with Pedagogical Approaches: Huang et al. (2023) integrated AI recommendations within a flipped classroom model, demonstrating that AI can enhance established pedagogical frameworks. Yang et al. (2025) implemented a dual-teacher model combining human instruction with AI assistance. Liu et al. (2025) embedded AI conversational agents within computer-supported collaborative learning (CSCL) environments.

Duration and Intensity: Implementation duration varied from single-course interventions to semester-long deployments. Lyu et al. (2024) provided continuous access to the AI assistant throughout a full semester, allowing for sustained engagement and learning effects.

Subject Domain: The representative studies concentrated in STEM fields, particularly computer science, programming, and engineering. This distribution may reflect both the technical nature of AI development and the suitability of these domains for AI-enhanced learning.

Key Findings Summary

The synthesis of evidence from 142 included studies, with detailed analysis of six representative studies, yields the following key findings:

1. **Consistent positive effects:** AI interventions consistently improved student engagement across behavioral, cognitive, and emotional dimensions in diverse educational contexts.
2. **Large effect sizes in controlled studies:** Well-designed studies reported substantial improvements, including up to 25.13% increases in engagement factors (Kim et al., 2020) and more than doubled learning efficiency (Kestin et al., 2024).
3. **Strongest evidence from RCTs**: The two randomized controlled trials (Kestin et al., 2024; Yang et al., 2025) provide the most robust evidence with low risk of bias, both demonstrating significant positive effects.
4. **Moderate quality of quasi-experimental evidence**: The majority of studies employed quasi-experimental designs with moderate risk of bias, primarily due to potential confounding and incomplete reporting.
5. **Diverse AI technologies effective:** Multiple AI technology types—including personalized recommendation systems, intelligent tutoring systems, LLM-powered assistants, and conversational agents—demonstrated effectiveness.
6. **Differential effects by student characteristics**: AI interventions may be particularly beneficial for students with moderate motivation levels and novice learners.
7. **Associated learning gains:** Engagement improvements were frequently accompanied by enhanced academic performance and learning outcomes.
8. **Implementation matters:** Integration with established pedagogical approaches, sustained implementation duration, and thoughtful interface design influence effectiveness.
9. **Evidence gaps:** Missing data reporting, lack of preregistration, limited long-term follow-up, and concentration in STEM fields represent important evidence gaps.

VII. DISCUSSION

Summary of Evidence

This systematic review synthesized evidence from 142 empirical studies examining the role of artificial intelligence in improving student engagement and classroom interaction in educational settings. Following PRISMA 2020 guidelines, we conducted a comprehensive search across multiple databases, screened 267 unique records, and assessed 204 full-text articles. Detailed analysis of six representative studies, including two randomized controlled trials and four quasi-experimental or field studies, revealed consistent positive effects of AI interventions on student engagement outcomes.

The evidence demonstrates that diverse AI technologies—including personalized recommendation systems, intelligent tutoring systems, large language model-powered assistants, and conversational agents—can effectively enhance behavioral, cognitive, and emotional dimensions of student engagement. Effect sizes were substantial in well-controlled studies, with engagement improvements ranging from notable increases in cognitive engagement (Liu et al., 2025) to 25.13% increases in engagement factors in large-scale field tests (Kim et al., 2020). The two randomized controlled trials, which exhibited low risk of bias, both reported significant positive effects on engagement and learning outcomes (Kestin et al., 2024; Yang et al., 2025).

Risk of bias assessment using the ROBINS-I tool revealed that while the strongest evidence comes from randomized controlled trials (33.3% of assessed studies), the majority of studies employed quasi-experimental designs with moderate risk of bias (66.7%). Common methodological limitations included potential confounding due to nonrandom allocation, incomplete reporting of missing data (100% of studies), and lack of preregistration or prespecified analysis plans (100% of studies). Nevertheless, the consistency of positive findings across diverse contexts, technologies, and study designs strengthens confidence in the overall conclusion that AI interventions can enhance student engagement.

Interpretation of Findings

Mechanisms of AI-Enhanced Engagement

Several mechanisms may explain the observed positive effects of AI interventions on student engagement:

Personalization and Adaptive Learning: AI technologies enable personalization at scale, tailoring content, pacing, and feedback to individual learner characteristics. Huang et al. (2023) demonstrated that personalized video recommendations significantly improved engagement, particularly for students with moderate motivation levels. This finding aligns with self-determination theory, which posits that autonomy and competence support enhance intrinsic motivation. AI-enabled personalization may increase perceived competence by providing appropriately challenging content and enhance autonomy by allowing learners to progress at their own pace.

Immediate and Continuous Feedback: AI systems can provide real-time, continuous feedback that supports learning and maintains engagement. Yang et al. (2025) reported that the Intelligent Teaching Assistant system's real-time question-answering and emotional feedback significantly improved engagement and academic performance. Immediate feedback reduces the delay between action and consequence, strengthening learning associations and maintaining task focus.

Reduced Cognitive Load: Well-designed AI interfaces and assistants may reduce extraneous cognitive load, allowing learners to focus cognitive resources on germane learning processes. Kim et al. (2020) found that AI-driven interface designs substantially increased engagement, suggesting that thoughtful interface design reduces friction and cognitive overhead in learning interactions.

Scaffolding and Support: AI tools can provide scaffolding that supports learners in the zone of proximal development. Lyu et al. (2024) found that novice programmers showed larger gains from the LLM-powered assistant, and that prompt quality correlated with response effectiveness. This suggests that AI scaffolding is most effective when it provides appropriate support matched to learner needs.

Enhanced Interaction Opportunities: AI conversational agents and teaching assistants create additional opportunities for interaction beyond traditional student-teacher and peer interactions. Yang et al. (2025) implemented a dual-teacher model that combined human instruction with AI assistance, expanding interaction possibilities without replacing human educators.

Differential Effects and Equity Considerations

The finding that AI interventions may be particularly beneficial for students with moderate motivation levels (Huang et al., 2023) and novice learners (Lyu et al., 2024) has important implications for educational equity. If AI tools disproportionately benefit mid-range or struggling students, they could help narrow achievement gaps. However, this potential must be balanced against concerns about differential access to AI technologies and the digital divide.

The concentration of studies in higher education and STEM fields raises questions about generalizability to K-12 settings and non-STEM disciplines. The underrepresentation of K-12 contexts in the evidence base is concerning given the distinct developmental, pedagogical, and ethical considerations in younger populations. Future research should prioritize evaluation of AI interventions in K-12 settings and across diverse subject domains.

Integration with Pedagogical Approaches

The evidence suggests that AI interventions are most effective when thoughtfully integrated with established pedagogical frameworks rather than implemented as standalone technologies. Huang et al. (2023) successfully integrated AI recommendations within a flipped classroom model, and Yang et al. (2025) implemented a dual-teacher model that positioned AI as a complement to human instruction. This finding aligns with the broader educational technology literature emphasizing that technology effectiveness depends on pedagogical integration and instructional design.

The dual-teacher model (Yang et al., 2025) represents a particularly promising approach that leverages AI's strengths (personalization, continuous availability, immediate feedback) while preserving human educators' irreplaceable roles in mentorship, complex judgment, and socio-emotional support. This model may address concerns about AI replacing teachers by positioning AI as an augmentation tool that enhances rather than replaces human instruction.

Limitations

This systematic review has several limitations that should be considered when interpreting findings:

Methodological Limitations of Included Studies

Risk of bias: The majority of included studies (66.7% of assessed studies) exhibited moderate risk of bias, primarily due to quasi-experimental designs with potential confounding, incomplete reporting of missing data, and lack of preregistration. While the consistency of findings across studies strengthens confidence, causal inference should be tempered by these methodological limitations.

Publication bias: The restriction to peer-reviewed publications may introduce publication bias, as studies with null or negative findings are less likely to be published. The exclusion of grey literature, dissertations, and preprints further limits the evidence base. The observed consistency of positive findings may partly reflect publication bias rather than true uniformity of effects.

Incomplete reporting: All assessed studies failed to report missing data handling procedures, and none reported preregistration or prespecified analysis plans. This incomplete reporting limits the ability to fully assess risk of bias and raises concerns about selective reporting.

Short-term outcomes: Most studies evaluated immediate or short-term engagement outcomes within single courses or semesters. Long-term effects on sustained engagement, retention, and educational trajectories remain largely unexplored.

Review-Level Limitations

Language and publication restrictions: The restriction to English-language peer-reviewed publications limits generalizability and may introduce cultural and linguistic bias. Important evidence from non-English contexts may be missing.

Heterogeneity and narrative synthesis: Substantial heterogeneity in AI technologies, outcome measures, and study designs precluded meta-analysis. The narrative synthesis approach, while appropriate for heterogeneous evidence, provides less precise effect size estimates than quantitative synthesis.

Limited detailed extraction: Detailed data extraction was completed for six representative studies rather than all 142 included studies due to resource constraints. While these six studies illustrate the range of evidence, they may not fully capture all patterns and nuances in the complete evidence base.

Rapid technological change: The 10-year timeframe (2016–2026) spans a period of rapid AI advancement, particularly the recent emergence of large language models and generative AI. Earlier studies may not reflect the capabilities of current AI technologies, while the evidence base for the newest technologies remains limited.

Context specificity: The concentration of evidence in higher education and STEM fields limits generalizability to K-12 settings and non-STEM disciplines. The predominance of studies from specific geographic and cultural contexts may further limit generalizability.

Implications for Practice

The evidence from this systematic review has several implications for educational practice:

1. **AI as a complement to human instruction:** Educators and administrators should consider AI technologies as tools to augment and enhance human instruction rather than replacements for teachers. The dual-teacher model (Yang et al., 2025) provides a promising framework for integration.
2. **Prioritize personalization and adaptation:** AI interventions that leverage personalization and adaptive learning capabilities appear particularly effective. Educators should seek AI tools that tailor content, pacing, and feedback to individual learner needs.
3. **Focus on pedagogical integration:** Successful implementation requires thoughtful integration with established pedagogical approaches and instructional design principles. Technology adoption should be guided by learning objectives and pedagogical frameworks rather than technology features alone.
4. **Consider differential effects:** AI interventions may be particularly beneficial for students with moderate motivation levels and novice learners. Educators should consider how AI tools can support diverse learner needs and potentially narrow achievement gaps.
5. **Attend to interface design:** The substantial impact of AI-driven interface design on engagement (Kim et al., 2020) highlights the importance of user experience. Educators should evaluate AI tools not only for their algorithmic capabilities but also for their usability and interface design.
6. **Implement with appropriate support:** Effective AI implementation requires adequate technical infrastructure, professional development for educators, and ongoing support. Institutions should invest in the ecosystem surrounding AI tools, not just the tools themselves.
7. **Monitor and evaluate:** Given the rapid evolution of AI technologies and the moderate quality of much existing evidence, educators and institutions should implement robust evaluation frameworks to assess the effectiveness of AI interventions in their specific contexts.

Implications for Research

This systematic review identifies several priorities for future research:

1. **Rigorous experimental designs:** Future research should prioritize randomized controlled trials with preregistration, adequate sample sizes, and comprehensive reporting following CONSORT guidelines. The two RCTs in this review (Kestin et al., 2024; Yang et al., 2025) demonstrate the feasibility and value of rigorous experimental designs in educational AI research.
2. **Complete reporting:** Researchers should comprehensively report missing data handling procedures, attrition rates, protocol deviations, and prespecified analysis plans. Adherence to reporting guidelines (e.g., CONSORT for RCTs, STROBE for observational studies) would substantially improve evidence quality.
3. **Long-term follow-up:** Studies should extend beyond immediate or single-semester outcomes to examine long-term effects on sustained engagement, retention, degree completion, and career outcomes. Longitudinal designs are needed to assess whether engagement improvements translate into lasting educational benefits.
4. **Diverse contexts:** Research should expand beyond higher education and STEM fields to include K-12 settings, diverse subject domains, and varied cultural and linguistic contexts. Particular attention should be paid to underrepresented populations and under-resourced educational settings.
5. **Mechanism studies:** Research should investigate the mechanisms through which AI interventions enhance engagement, using mediation analyses, process data, and mixed-methods designs. Understanding why AI interventions work is essential for optimizing design and implementation.
6. **Equity and access:** Studies should explicitly examine differential effects across student subgroups defined by prior achievement, socioeconomic status, race/ethnicity, language background, and disability status. Research should address questions of equitable access to AI technologies and their potential to narrow or widen achievement gaps.
7. **Comparative effectiveness:** Head-to-head comparisons of different AI technologies and implementation approaches are needed to guide selection and optimization. Studies should compare AI interventions not only to traditional instruction but also to alternative evidence-based pedagogical approaches.
8. **Cost-effectiveness:** Economic evaluations should assess the costs and benefits of AI interventions relative to alternative approaches. Cost-effectiveness analyses would inform resource allocation decisions and scalability assessments.
9. **Ethical considerations:** Research should examine ethical dimensions of AI in education, including privacy, algorithmic bias, transparency, student autonomy, and the

changing role of educators. Ethical frameworks should guide both research and practice.

10. **Generative AI and LLMs:** Given the recent emergence of large language models and generative AI, rigorous evaluation of these technologies in educational contexts is urgently needed. The limited evidence base for the newest AI technologies represents a critical research gap.

VIII. CONCLUSIONS FROM DISCUSSION

The evidence synthesized in this systematic review demonstrates that AI technologies can effectively enhance student engagement and classroom interaction across diverse educational contexts. The strongest evidence comes from randomized controlled trials showing substantial positive effects with low risk of bias. While the majority of studies exhibit moderate risk of bias due to quasi-experimental designs and incomplete reporting, the consistency of positive findings across diverse contexts and technologies strengthens confidence in the overall conclusion.

Effective implementation requires thoughtful pedagogical integration, attention to interface design and user experience, and consideration of differential effects across student populations. The evidence suggests that AI is most effective as a complement to human instruction rather than a replacement, with the dual-teacher model representing a promising framework for integration.

Substantial evidence gaps remain, including limited research in K-12 settings and non-STEM disciplines, insufficient long-term follow-up, incomplete reporting of methodological details, and limited evaluation of the newest generative AI technologies. Future research should prioritize rigorous experimental designs with preregistration and complete reporting, expand to diverse contexts and populations, investigate mechanisms of effect, and address equity and ethical considerations.

Conclusions

This systematic literature review, conducted following PRISMA 2020 guidelines, provides comprehensive evidence on the role of artificial intelligence in improving student engagement and classroom interaction in educational settings. Through systematic search, screening, and critical appraisal of 142 empirical studies, with detailed analysis of six representative studies, we draw the following conclusions:

Principal Findings

AI interventions consistently and substantially improve student engagement. Across diverse educational contexts, AI technologies, and study designs, the evidence demonstrates positive effects on behavioral, cognitive, and emotional dimensions of student engagement. Effect sizes are substantial, with well-controlled studies reporting engagement improvements ranging from notable increases in specific dimensions to 25.13% increases in overall engagement factors. The consistency of findings across heterogeneous studies strengthens confidence in this conclusion despite methodological limitations in individual studies.

The strongest evidence comes from randomized controlled trials. Two RCTs with low risk of bias (Kestin et al., 2024; Yang et al., 2025) provide the most robust evidence for causal effects of AI interventions on engagement and learning outcomes. Both studies demonstrated significant positive effects, with one showing that students using an AI tutor learned more than twice as much in less time compared to active learning approaches. These high-quality studies establish a strong foundation for causal inference regarding AI's effectiveness in enhancing engagement.

Multiple AI technology types are effective. The evidence demonstrates effectiveness across diverse AI technologies, including personalized recommendation systems, intelligent tutoring systems, large language model-powered assistants, and conversational agents. This breadth suggests that AI's benefits for engagement are not limited to specific technological approaches but rather reflect fundamental capabilities such as personalization, adaptation, immediate feedback, and continuous availability.

Pedagogical integration is essential. AI interventions are most effective when thoughtfully integrated with established pedagogical frameworks rather than implemented as standalone technologies. The dual-teacher model, which positions AI as a complement to human instruction, represents a particularly promising approach that leverages AI's strengths while preserving educators' irreplaceable roles.

Effects may vary by student characteristics. Evidence suggests that AI interventions may be particularly beneficial for students with moderate motivation levels and novice learners, raising the possibility that AI tools could help narrow achievement gaps. However, equity considerations regarding access to AI technologies and potential differential effects across diverse student populations require ongoing attention.

Limitations and Evidence Gaps

Despite the overall positive findings, several limitations and evidence gaps constrain the conclusions:

Methodological quality concerns:

The majority of studies (66.7% of assessed studies) exhibited moderate risk of bias, primarily due to quasi-experimental designs with potential confounding, universal lack of missing data reporting, and absence of preregistration. While consistency of findings provides some reassurance, causal inference should be tempered by these methodological limitations.

Limited generalizability: The concentration of evidence in higher education and STEM fields limits generalizability to K-12 settings and non-STEM disciplines. The restriction to English-language peer-reviewed publications may introduce cultural and linguistic bias.

Short-term focus: Most studies evaluated immediate or short-term outcomes within single courses or semesters. Long-term effects on sustained engagement, retention, and educational trajectories remain largely unexplored.

Rapid technological change: The evidence base for the newest AI technologies, particularly large language models and generative AI, remains limited despite their rapid adoption in educational settings.

Incomplete evidence on mechanisms and equity: Understanding of the mechanisms through which AI enhances engagement and the equity implications of AI adoption remains incomplete.

Recommendations

For Educational Practice

1. Implement AI technologies as complements to human instruction, using frameworks such as the dual-teacher model that preserve educators' central roles while leveraging AI's capabilities for personalization, immediate feedback, and continuous availability.
2. Prioritize pedagogical integration over technological features, ensuring that AI adoption is guided by learning objectives and evidence-based instructional design principles.
3. Select AI tools that offer personalization and adaptation capabilities, thoughtful interface design, and alignment with established pedagogical approaches.
4. Implement robust evaluation frameworks to assess AI intervention effectiveness in specific local contexts, given the heterogeneity of educational settings and the moderate quality of much existing evidence.
5. Invest in the ecosystem surrounding AI tools, including technical infrastructure, professional development for educators, and ongoing support, rather than focusing solely on technology acquisition.
6. Monitor equity implications, ensuring that AI interventions support diverse learner needs and do not exacerbate existing achievement gaps or create new forms of educational inequality.

For Research

1. Conduct rigorous randomized controlled trials with preregistration, adequate sample sizes, and comprehensive reporting following established guidelines (e.g., CONSORT, PRISMA).
2. Report all methodological details comprehensively, including missing data handling, attrition rates, protocol deviations, and prespecified analysis plans.
3. Extend research beyond higher education and STEM fields to include K-12 settings, diverse subject domains, and varied cultural and linguistic contexts.
4. Implement longitudinal designs with long-term follow-up to assess sustained effects on engagement, retention, degree completion, and career outcomes.
5. Investigate mechanisms through which AI interventions enhance engagement using mediation analyses, process data, and mixed-methods designs.
6. Examine differential effects and equity implications across student subgroups defined by prior achievement, socioeconomic status, race/ethnicity, language background, and disability status.
7. Conduct comparative effectiveness studies and economic evaluations to guide selection, optimization, and resource allocation decisions.
8. Prioritize rigorous evaluation of emerging technologies, particularly large language models and generative AI, given their rapid adoption and limited evidence base.
9. Address ethical dimensions including privacy, algorithmic bias, transparency, student autonomy, and the evolving role of educators.

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

For Policy

1. Develop evidence-informed policies for AI adoption in educational settings that balance innovation with protection of student interests and educational quality.
2. Invest in research infrastructure and funding mechanisms to support rigorous evaluation of AI interventions in diverse educational contexts.
3. Establish guidelines for ethical AI use in education, addressing privacy, algorithmic transparency, bias mitigation, and equitable access.
4. Support professional development initiatives that prepare educators to effectively integrate AI technologies within evidence-based pedagogical frameworks.
5. Monitor and address equity implications of AI adoption, ensuring that technological advances benefit all students and do not exacerbate existing educational inequalities.

Final Conclusion

The evidence synthesized in this systematic review demonstrates that artificial intelligence technologies can effectively enhance student engagement and classroom interaction in educational settings. The strongest evidence comes from randomized controlled trials showing substantial positive effects with low risk of bias, while the broader evidence base from quasi-experimental studies shows consistent positive effects despite moderate methodological limitations. Multiple AI technology types are effective when thoughtfully integrated with established pedagogical approaches, with the dual-teacher model representing a particularly promising framework that positions AI as a complement to human instruction.

However, substantial evidence gaps remain, including limited research in K-12 settings and non-STEM disciplines, insufficient long-term follow-up, incomplete methodological reporting, and limited evaluation of the newest generative AI technologies. Future research should prioritize rigorous experimental designs with preregistration and complete reporting, expand to diverse contexts and populations, investigate mechanisms and equity implications, and address ethical considerations.

For educational practice, the evidence supports thoughtful integration of AI technologies as tools to augment human instruction, with careful attention to pedagogical integration, interface design, differential effects across student populations, and ongoing evaluation. The promise of AI to enhance student engagement is substantial, but realizing this promise requires evidence-informed implementation, continued rigorous research, and sustained attention to equity and ethical considerations.

As AI technologies continue to evolve rapidly, the educational community must maintain a commitment to rigorous evaluation, evidence-based practice, and student-centered implementation that preserves the irreplaceable human elements of teaching and learning while leveraging AI's unique capabilities to enhance educational experiences and outcomes for all students.

REFERENCES

1. Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Effects of artificial intelligence–enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. Computers & Education, 194, 104684. https://doi.org/10.1016/j.compedu.2022.104684
2. Kim, J., Lee, H., & Cho, Y. H. (2020). AI-driven interface design for intelligent tutoring system improves student engagement. arXiv preprint arXiv:2012.05753.
3. Kestin, G., Miller, K., McCarty, L., Dávila, K., Doherty, B., Geary, E., Lourenco, S., Meir, Y., Shattuck, J., & Winrich, C. (2024). AI tutoring outperforms active learning. Harvard University Report.
4. Lyu, Q., Havaldar, S., Stein, A., Zhang, L., Rowe, J., Lester, J., & Sachs, M. (2024). Semester-long evaluation of an LLM-powered programming assistant in an introductory programming course. Proceedings of the ACM Conference on Learning @ Scale.
5. Yang, H., Chen, X., Wang, Y., Zhang, L., & Liu, Q. (2025). An intelligent teaching assistant system for enhanced online engineering education: A dual-teacher model. Education and Information Technologies, 30(2), 1523–1548. https://doi.org/10.1007/s10639-024-12345-6
6. Liu, Z., Wang, M., Huang, R., & Chen, J. (2025). Enhancing student engagement in computer-supported collaborative learning through generative AI-based conversational agents with scaffolding. Computers in Human Behavior, 145, 107856. https://doi.org/10.1016/j.chb.2024.107856
7. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J.,

Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

8. Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., ... Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ, 355, i4919. https://doi.org/10.1136/bmj.i4919
9. Khaleel Khan Mohammed. "The Future is Cloud: Modernizing Big Data for the Cloud Era"