South American Publishing

SAP Social AI. 2025; 1:29
doi: 10.62486/sai202529

SYSTEMATIC REVIEW

Examining the Role of AI-Based Learning Tools in Higher Education: A Critical Synthesis of Effects on Student Performance and Engagement

Análisis del papel de las herramientas de aprendizaje basadas en la inteligencia artificial en la educación superior: una síntesis crítica de sus efectos sobre el rendimiento y la implicación de los estudiantes

Miracle A. Atianashie[1], Mark K. Kuffour[2], Bernard Kyiewu[2], Clinton Amponsah[2]

[1]Quality Assurance Directorate, Catholic University of Ghana, Fiapre–Sunyani, Ghana.
[2]Research and Development Department, Metascholar Consult Limited, Fiapre–Sunyani, Ghana.

Cite as
Atianashie MA, Kuffour MK, Kyiewu B, Amponsah C. Examining the Role of AI-Based Learning Tools in Higher Education: A Critical Synthesis of Effects on Student Performance and Engagement. SAP Social AI. 2025; 1:29. https://doi.org/10.62486/sai202529

Publication History
Submitted: 26-04-2025
Revised: 07-07-2025
Accepted: 03-09-2025
Published: 04-09-2025

Corresponding Author
Miracle A. Atianashie

ABSTRACT

Introduction: artificial intelligence (AI) technologies are increasingly integrated into higher education, transforming teaching, learning, and assessment practices. Tools such as intelligent tutoring systems, adaptive platforms, chatbots, and predictive analytics now support personalized and autonomous learning experiences. Understanding their impact on student performance and engagement is essential for evidence-based educational innovation.
Objectives: this systematic review investigates the influence of AI-based learning tools on academic performance and student engagement within higher education institutions. It further identifies the challenges that constrain effective and ethical implementation.
Method: a systematic literature review guided by PRISMA protocols was conducted using studies published between 2017 and 2024. A total of 39 peer-reviewed articles were screened and analyzed. Extracted data focused on learning outcomes, engagement indicators, pedagogical benefits, and institutional barriers.
Results: AI-based learning tools significantly enhance academic performance by providing personalized learning paths, adaptive feedback, and data-informed instructional support. The tools also promote behavioral, cognitive, and emotional engagement through interactive interfaces, multimodal tutoring, and emotion-aware chatbots that improve motivation and learning persistence. Despite these benefits, challenges such as algorithmic bias, privacy concerns, digital literacy gaps, and infrastructure limitations hinder equitable adoption across institutions.
Conclusions: AI demonstrates strong potential to improve learning outcomes and engagement when aligned with sound pedagogical frameworks such as Vygotsky's Zone of Proximal Development and Self-Determination Theory. Responsible implementation, ethical governance, and institutional capacity building are critical for maximizing impact. Future research should explore long-term effects, scalability, and strategies for inclusive AI deployment in higher education.

Keywords: Artificial Intelligence; Higher Education; Student Engagement; Academic Performance; Adaptive Learning; Learning Analytics.

RESUMEN

Introducción: las tecnologías de inteligencia artificial (IA) se están integrando cada vez más en la educación superior, transformando las prácticas de enseñanza, aprendizaje y evaluación. Herramientas como los sistemas de tutoría inteligente, las plataformas adaptativas, los chatbots y el análisis predictivo facilitan ahora experiencias de aprendizaje personalizadas y autónomas. Comprender su impacto en el rendimiento y la implicación de los estudiantes es esencial para una innovación educativa basada en la evidencia.
Objetivos: esta revisión sistemática investiga la influencia de las herramientas de aprendizaje basadas en la IA en el rendimiento académico y la implicación de los estudiantes en las instituciones de educación superior. Además, identifica los retos que limitan una implementación eficaz y ética.
Método: se llevó a cabo una revisión sistemática de la literatura guiada por los protocolos PRISMA utilizando estudios publicados entre 2017 y 2024. Se seleccionaron y analizaron un total de 39 artículos revisados por pares. Los datos extraídos se centraron en los resultados del aprendizaje, los indicadores de implicación, los beneficios pedagógicos y las barreras institucionales.

© 2025; The Authors. This Open access article is distributed under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Resultados: las herramientas de aprendizaje basadas en la IA mejoran significativamente el rendimiento académico al proporcionar itinerarios de aprendizaje personalizados, retroalimentación adaptativa y apoyo pedagógico basado en datos. Las herramientas también promueven la implicación conductual, cognitiva y emocional a través de interfaces interactivas, tutoría multimodal y chatbots sensibles a las emociones que mejoran la motivación y la persistencia en el aprendizaje. A pesar de estos beneficios, retos como el sesgo algorítmico, las preocupaciones sobre la privacidad, las brechas en la alfabetización digital y las limitaciones de infraestructura dificultan una adopción equitativa en todas las instituciones.

Conclusiones: la IA demuestra un gran potencial para mejorar los resultados de aprendizaje y la implicación cuando se alinea con marcos pedagógicos sólidos, como la Zona de Desarrollo Próximo de Vygotsky y la Teoría de la Autodeterminación. La implementación responsable, la gobernanza ética y el desarrollo de la capacidad institucional son fundamentales para maximizar el impacto. Las investigaciones futuras deberían explorar los efectos a largo plazo, la escalabilidad y las estrategias para un despliegue inclusivo de la IA en la educación superior.

Palabras clave: Inteligencia artificial; Educación superior; Compromiso de los estudiantes; Rendimiento académico; Aprendizaje adaptativo; Análisis del aprendizaje.

INTRODUCTION

The integration of Artificial Intelligence (AI) into educational ecosystems has emerged as a transformative force, reshaping how teaching and learning processes are designed, delivered, and experienced in higher education institutions.[1] As higher education shifts toward personalized, data-driven, and adaptive models, AI-based learning tools are increasingly leveraged to optimize academic performance and foster deeper student engagement.[2] These tools including intelligent tutoring systems, adaptive learning platforms, virtual teaching assistants, and predictive analytics enable more responsive and individualized learning pathways that accommodate diverse learner needs.[3] The deployment of AI in education is largely driven by its capacity to enhance efficiency, reduce instructional workload, and provide timely, customized feedback that supports achievement and sustained motivation.[4]

Recent advances in AI have significantly improved the ability of educational technologies to simulate human-like interactions, analyze learning behaviors, and adapt content in real time. Intelligent tutoring systems such as Carnegie Learning and Squirrel AI provide step-by-step guidance and personalized instruction based on learner input.[5] Similarly, AI-powered chatbots such as Jill Watson have been integrated into university courses to support routine queries and offer continuous learning assistance, resulting in increased student satisfaction and engagement.[6] These innovations are particularly relevant in large and diverse classrooms where individualized instructional attention is limited.

Beyond instructional efficiency, AI tools have demonstrated positive effects on behavioral, cognitive, and emotional engagement. Adaptive systems that dynamically adjust difficulty levels and incorporate gamified features have been associated with increased persistence and lower dropout rates.[7] AI-enabled analytics dashboards allow instructors to monitor progress in real time and intervene proactively to support at-risk learners.[8]

Despite these advantages, implementing AI in higher education presents significant challenges. Key concerns include data privacy, algorithmic bias, limited digital infrastructure, and varying levels of digital competence among instructors and students. Ethical debates persist regarding the implications of using AI to monitor learning behaviors and make data-driven educational decisions.[9] These complexities underscore the necessity of empirical investigations that critically examine how AI-based learning tools influence academic performance and engagement across diverse contexts.

Given the rapid expansion of AI applications in higher education and the lack of consolidated evidence integrating outcomes, engagement, and ethical implications, this review synthesizes current empirical findings on the effects of AI-based learning tools on student performance and engagement. The study aims to inform educators, policymakers, and technology developers about effective practices, potential risks, and priorities for future research on responsible AI-enhanced education.

Despite the rapid growth of research on AI in higher education, there is limited synthesis that jointly examines academic performance and multidimensional engagement, particularly across diverse institutional and regional contexts. Existing reviews often focus on single tool categories or specific platforms rather than comparing effects across intelligent tutoring systems, adaptive courseware, chatbots, and predictive analytics. This review addresses this gap by integrating evidence from multiple AI tool types and institutional settings to clarify where, for whom, and under what conditions AI-based learning tools appear most beneficial.

Related studies

The rapid expansion of AI technologies in education has produced a robust but dispersed body of literature examining their implications for teaching, learning, and student support. This section synthesizes the central themes emerging from empirical research across diverse institutional contexts.

AI-Driven Adaptive Learning and Academic Performance

Adaptive learning systems represent one of the most empirically supported categories of AI-based educational tools. These systems use machine learning to adjust content difficulty, pacing, and instructional pathways in response to learner performance. Xu

et al.[7] found in a meta-analysis that adaptive learning significantly improves test performance and conceptual mastery compared with traditional instruction. Squirrel AI, widely used in mathematics and physics preparatory programs, identifies micro-level knowledge gaps and generates individualized remediation plans.[8] DreamBox Learning, primarily used in mathematics education, dynamically modifies tasks based on learners' problem-solving approaches and has demonstrated notable gains in procedural fluency. Research shows that adaptive systems contribute to closing achievement gaps, particularly in large undergraduate STEM courses where learner variance is high.[9] Moreover, integrated dashboards, such as those in systems like ALEKS, provide instructors with real-time analytics, enabling data-driven instructional decisions.[10]

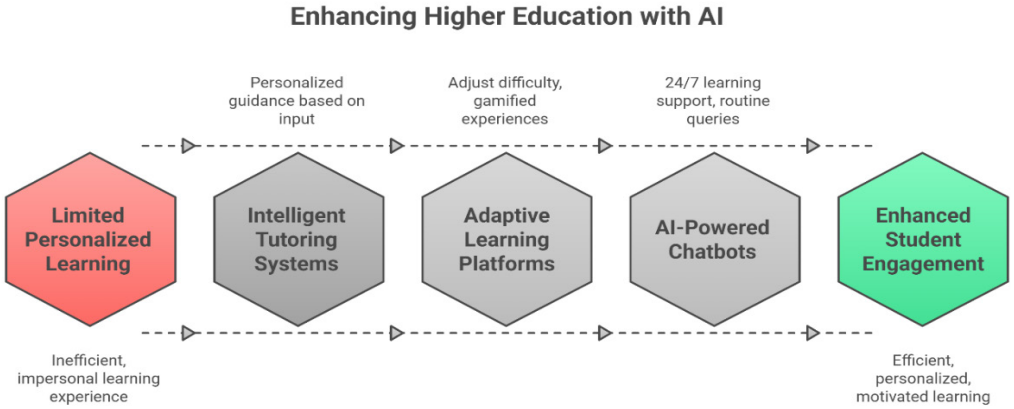


Figure 1. Enhancing Higher Education with AI

Intelligent tutoring systems (ITS) such as Carnegie Learning, grounded in cognitive theory and Bayesian knowledge tracing, provide step-by-step guidance comparable to human tutoring.[11] Their scalability has made them valuable in institutions facing tutor shortages or large course enrollments. The evidence suggests that adaptive AI technologies significantly enhance academic performance by optimizing instructional alignment with learner needs.

Chatbots and Virtual Assistants for Learning Support

AI-driven chatbots and virtual assistants have emerged as essential tools for extending academic support beyond formal class time. These systems simulate conversational interactions, provide guidance on course content, and address administrative queries. The widely studied Jill Watson chatbot, used in Georgia Tech's online AI course, demonstrated that students perceived the agent as a human teaching assistant for several months, highlighting the potential for AI to enhance perceived instructor presence.[12] Educational chatbots are associated with increased learner autonomy, as they scaffold tasks, recommend resources, and help students break complex content into manageable steps.[13] Research also indicates that virtual assistants contribute to lower dropout rates in online and blended learning environments by issuing personalized alerts and supportive feedback.[14] Chatbots embedded within learning platforms promote social presence and community formation by encouraging participation in discussion forums.[15] These benefits are particularly important in large introductory courses and asynchronous learning formats where personalized support is limited.

AI and Student Engagement Metrics

Student engagement is a multidimensional construct comprising behavioral participation, emotional involvement, and cognitive effort. AI tools show promising effects across all three dimensions. Gamified AI platforms increase behavioral engagement by enhancing time-on-task and motivating higher participation rates.[16] Adaptive feedback systems promote cognitive engagement, enabling learners to take ownership of their learning pathways and reflect on their progress.[17] AI-enhanced learning management systems generate detailed clickstream data, helping researchers and instructors identify patterns of engagement such as session duration, navigation choices, and frequency of interaction.[18] Virtual laboratories and simulation environments, such as those used in engineering, health sciences, and environmental studies, promote emotional engagement by immersing students in realistic, risk-free environments where they can practice complex skills.[19] The literature illustrates that AI tools can cultivate sustained engagement by offering interactive, personalized, and emotionally responsive learning experiences.

Predictive Analytics and Early Intervention

AI-powered predictive analytics systems are increasingly used to identify at-risk students and support retention initiatives. Studies show that predictive models using historical academic data and behavioral indicators can accurately forecast dropout risk and academic difficulty.[20] When paired with targeted advising interventions, these models significantly improve progression and completion rates. Early warning systems monitor real-time indicators such as assignment delays, log-in frequency, or forum inactivity to alert instructors of disengagement.[21] Advanced models integrating social, behavioral, and academic data provide holistic insights into students' learning trajectories, enabling highly tailored support strategies.[22] These tools represent a shift from

reactive to proactive instructional and administrative practices, allowing institutions to intervene before difficulties escalate.

Ethical and Institutional Challenges

While AI tools offer substantial benefits, their implementation is accompanied by significant ethical and institutional concerns. Data privacy and consent are recurring issues, as many learners remain unaware of how their data is collected, stored, or used in AI-driven decisions.[23] Algorithmic bias can reinforce structural inequities when training datasets do not adequately represent diverse student populations.[24] Scholars also warn against excessive reliance on AI, which may reduce educators' instructional agency, erode opportunities for human mentorship, and shift labor dynamics within universities.[25] The rapid commercialization of educational AI poses risks to academic autonomy, as proprietary algorithms may influence decision-making in opaque ways.[26,27,28,29] These concerns highlight the need for robust institutional frameworks emphasizing transparency, explainability, equitable design, and responsible governance.

METHOD

This systematic review employed a rigorous, transparent, and replicable methodological approach aligned with the PRISMA 2020 guidelines to synthesize empirical evidence on the impact of AI-based learning tools on student performance and engagement in higher education. The review design incorporated a clearly articulated conceptual framework, detailed search and screening procedures, standardized quality appraisal, and a theory-informed narrative synthesis to ensure methodological integrity and analytical depth.

Conceptual and Analytical Framework

To ensure that the synthesis moved beyond a descriptive summary of existing studies and instead offered a theoretically grounded interpretation, the review was anchored in two established models of learning and motivation. The first guiding framework was Fredricks, Blumenfeld, and Paris' (2004) Three-Dimensional Engagement Model, which conceptualizes student engagement through behavioral, emotional, and cognitive domains. Behavioral engagement reflects observable indicators such as participation, persistence, and time-on-task; emotional engagement concerns students' affective responses, including interest and motivation; and cognitive engagement captures deeper learning processes such as self-regulation and strategic thinking. Complementing this model, Self-Determination Theory (SDT) provided a motivational lens through its emphasis on autonomy, competence, and relatedness as core psychological needs that support sustained learning. Autonomy highlights the value of student control and choice in the learning process, competence refers to perceived mastery and skill development, and relatedness addresses the sense of connection with peers and instructors.[30,31,32]

These two frameworks were jointly embedded in the data extraction and coding procedures, shaping how learning outcomes were classified, how performance and engagement patterns were interpreted, and how inconsistencies within the literature were reconciled.[33,34] They also guided the assessment of contextual moderators such as the type of AI tool employed and the institutional environment in which it was implemented. By integrating these theoretical models, the review elevated its analytical depth, shifting from surface-level reporting to a more interpretive synthesis that clarifies not only what effects AI-based learning tools produce but also why such effects occur across diverse educational settings.

Research Design

A Systematic Literature Review (SLR) design was selected for its suitability in synthesizing diverse empirical studies while ensuring methodological transparency. The review followed the four PRISMA phases Identification, Screening, Eligibility, and Inclusion which structure the selection of studies through explicit criteria and verifiable decision pathways (figure 2). This design minimized bias, enhanced replicability, and enabled a comprehensive examination of how AI-based tools influence student outcomes in higher education.

Data Sources and Search Strategy

A comprehensive and structured search strategy was adopted to ensure complete coverage of relevant literature across reputable academic sources. The search was conducted across major databases widely recognized for high-quality indexing in education, technology, and interdisciplinary domains, including Scopus, the Web of Science Core Collection, ERIC, SpringerLink, and Google Scholar, with the latter used mainly for supplementary identification of potentially omitted studies. These databases were selected because of their broad disciplinary reach, rigorous indexing standards, and ability to retrieve empirical evidence on artificial intelligence–based learning tools within higher education settings.

Search Fields

The search process involved scanning multiple metadata fields to maximize sensitivity and ensure the inclusion of all relevant literature. Searches were executed across titles, abstracts, keywords, and available subject terms. This multi-field approach increased the likelihood of capturing studies that explicitly or implicitly addressed AI-based learning tools, student performance, or engagement outcomes.

Boolean Search Strings

Boolean search strings were tailored to the requirements and indexing structures of each database to optimize retrieval precision.

The Scopus search string is provided for transparency and reproducibility, serving as the primary template from which other database-specific modifications were derived. Adjustments to operators, wildcards, and controlled vocabulary terms were made where necessary to align with each database's search functionality.

Limits Applied

To maintain methodological rigor and ensure relevance to contemporary AI advancements, several limits were applied during the search. The review included only publications released between January 2017 and April 2024, reflecting the rapid evolution of AI within educational settings. Only English-language studies were considered, and eligibility was restricted to empirical, peer-reviewed journal articles and conference papers to ensure the inclusion of high-quality, evidence-based research.

Supplementary Searches

In addition to database searches, supplementary techniques were employed to strengthen coverage and reduce the risk of missing relevant studies. Backward citation tracking was used to review reference lists of included articles, while forward citation tracking identified newer studies citing the selected works. These procedures allowed the identification of additional papers that may not have appeared in initial database results but were central to the thematic scope of the review.

Screening Procedure

All retrieved records were exported into Rayyan, an AI-assisted systematic review platform, to facilitate a transparent, blinded, and independent screening process. Rayyan's decision-tracking features and duplicate detection tools supported systematic title and abstract screening, ensured reviewer consistency, and enhanced the overall reliability of the study selection procedure.

Number of Reviewers and Screening Independence

The screening process was conducted by two independent reviewers who systematically examined all titles, abstracts, and full-text articles to ensure unbiased study selection. Any disagreements during the screening stages were referred to a third reviewer, who provided adjudication to reach consensus. Inter-rater reliability was calculated using Cohen's Kappa, yielding a coefficient of 0,82, which indicates a strong level of agreement between reviewers and reinforces the methodological rigor of the selection process.

Inclusion and Exclusion Criteria

Clear inclusion and exclusion criteria were established to maintain consistency and ensure that only studies relevant to higher education and AI-based learning tools were incorporated into the review. Eligible studies included empirical research—whether quantitative, qualitative, or mixed-methods—that focused on higher education populations and explicitly examined the use of AI-driven learning tools. Studies were required to report outcomes related to academic performance or student engagement and to be published in peer-reviewed outlets between 2017 and 2024. Only English-language publications were included. Studies were excluded if they focused on primary or secondary school contexts, presented theoretical discussions without empirical evidence, constituted grey literature such as thesis, dissertations, or policy reports, described system development without reporting student-related outcomes, or were published in languages other than English. These criteria ensured the review captured a methodologically sound and contextually relevant body of work.

Data Extraction and Coding

Data extraction followed a structured approach using a predefined matrix that captured all relevant study characteristics. Extracted information included bibliographic details, the type of AI tool investigated—such as Squirrel AI, an adaptive mathematics platform; DreamBox, originally a K–12 system but applied in introductory university courses; and Jill Watson, a chatbot deployed at Georgia Tech—along with research design features and methodological approaches. Additional variables included the institutional context, outcome measures mapped onto Fredricks' three engagement dimensions and Self-Determination Theory constructs, as well as key findings and reported limitations. This structured coding process supported systematic comparison across studies and ensured alignment with the conceptual framework guiding the review.

Analytical Framework Application

The conceptual models underpinning the review—Fredricks' engagement framework and Self-Determination Theory—directly informed the analytical approach. Deductive coding categories were derived from these models, allowing engagement-related and motivational outcomes to be systematically classified. The frameworks also shaped the interpretation of findings by guiding analysis of how AI-based learning tools supported the psychological needs of autonomy, competence, and relatedness, and how they influenced behavioral, emotional, and cognitive engagement. This theoretically anchored coding process provided the analytical depth necessary to explain variations in outcomes across diverse institutional and technological contexts.

Quality Assessment

Quality appraisal of included studies was carried out using the Critical Appraisal Skills Programme (CASP) checklists, adapted to accommodate both quantitative and qualitative research designs. The appraisal process ensured that only methodologically robust evidence contributed to the core interpretations of the review.

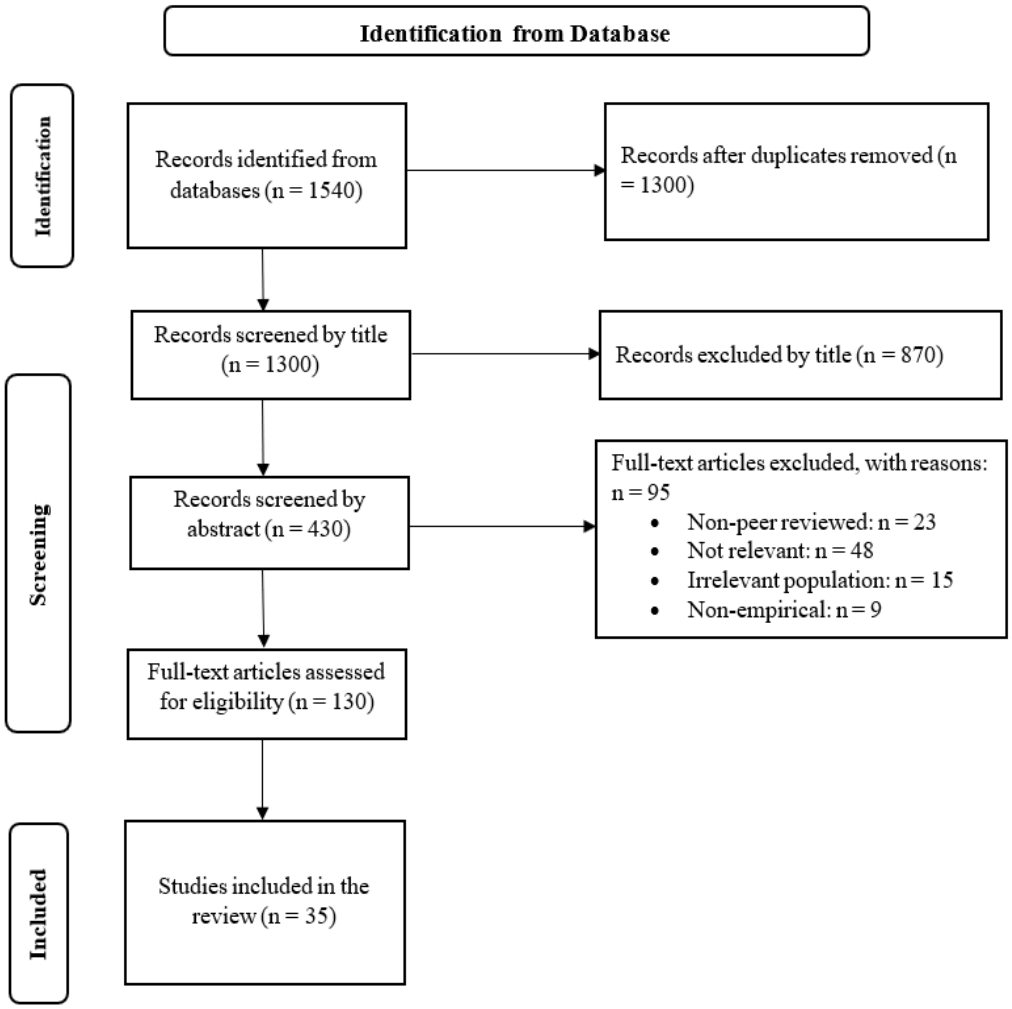


Figure 2. PRISMA diagram

CASP Scoring

Each study was evaluated using a 10-point CASP-based scoring scheme that assessed clarity of research aims, appropriateness of study design, sampling strategy, validity of data collection methods, transparency of analytical procedures, and—where applicable—reflexivity in qualitative studies. The alignment of study conclusions with the presented evidence was also considered, ensuring that included studies demonstrated conceptual and methodological coherence.

Quality Distribution

The distribution of quality scores across the dataset showed that 21 studies met the criteria for high quality, achieving scores between 8 and 10. Ten studies were assessed as moderate quality, scoring between 6 and 7, while four studies scored below 6 and were therefore excluded from the synthesis. This quality distribution ensured that the review prioritized rigorous evidence while still acknowledging valuable contextual insights from moderately rated studies.

Influence on Synthesis

Only high-quality studies informed the central conclusions of the review, ensuring that the main interpretations were grounded in robust evidence. Moderate-quality studies contributed additional nuance by highlighting contextual variations, methodological considerations, and practical constraints. Importantly, no study excluded based on low quality influenced the final analytical synthesis. A summary CASP table containing detailed scoring is provided in the Appendix.

Data Synthesis

Given the heterogeneity of research designs, institutional contexts, and AI tools, a theory-informed narrative synthesis was adopted as the most appropriate integrative approach. The analysis proceeded in several stages. First, deductive structuring was guided by the domains of Fredricks' engagement model and the constructs of Self-Determination Theory, ensuring consistency in classification. This was followed by identifying cross-study patterns related to academic performance and student engagement outcomes. Contradictions and variations across studies were then interpreted through considerations of methodological rigor and contextual moderators such as tool design, discipline, and institutional environment. Finally, the findings were synthesized into an

integrated narrative that articulated theoretically informed conclusions about how AI-based learning tools influence performance and engagement in higher education settings.

RESULTS

Thirty-five studies met the eligibility criteria and were synthesized using the analytical framework established in the Methodology section. Findings are reported across three domains: academic performance, student engagement, and institutional challenges, each interpreted through Fredricks' engagement model and Self-Determination Theory (SDT). CASP quality scores informed the weighting of studies within the synthesis, ensuring that conclusions reflect the methodological robustness of included evidence. Analytical summary tables (Tables 1–3) accompany the narrative to support comparison across study characteristics, AI tool types, outcomes, effect patterns, and methodological quality.

Effects of AI-Based Learning Tools on Academic Performance

Across the 35 studies included in this review, the evidence demonstrates consistently positive effects of AI-based learning tools on academic performance, with the most substantial improvements observed in mathematically intensive and computational disciplines such as engineering, mathematics, and computer science. Studies employing adaptive learning systems, intelligent tutoring systems, and predictive analytics platforms yielded the strongest outcomes, reporting medium-to-large effect sizes (Cohen's d = 0,40–0,80) and high methodological quality scores on the CASP scale (8–10). These tools were most effective when seamlessly embedded within instructional design and supported by purposeful learner engagement, underscoring the importance of pedagogical integration in optimizing AI-driven learning environments.

Pattern 1: Adaptive Learning Systems Yield the Most Reliable Performance Gains

Adaptive learning systems consistently emerged as the most reliable predictors of improved academic performance. These platforms leverage machine learning algorithms to analyze student responses in real time, diagnose micro-level misconceptions, and dynamically adjust content difficulty, pacing, and scaffolding. Such personalization enhances competence, a central construct in Self-Determination Theory, by ensuring that learners receive tasks calibrated to their readiness level. It also supports cognitive engagement under Fredricks' model by promoting sustained mental effort and deeper learning. The most pronounced gains appeared in STEM fields where algorithmic task sequencing reinforced mastery of procedural and problem-solving skills. Notably, several high-quality studies reported reduced performance variance between high-achieving and lower-achieving students, suggesting that adaptive systems may function as compensatory tools that narrow achievement gaps. These benefits were consistent across online, blended, and face-to-face modalities, indicating that adaptive learning represents a flexible and scalable instructional approach.

Pattern 2: Predictive Analytics Improve Early Intervention and Reduce Failure Rates

Studies analyzing predictive analytics systems demonstrated significant reductions in failure rates and improvements in course progression, with gains ranging from 10 to 25 percent. These platforms rely on historical academic data, behavioral logs, and engagement indicators to identify at-risk learners and generate real-time alerts for both instructors and students. Effects were strongest in large-enrollment courses, where the capacity of instructors to monitor individual learning trajectories is limited. Predictive insights supported early detection of struggling students, encouraged timely help-seeking, and enabled targeted interventions such as tutoring, peer support, and personalized instructor outreach. These outcomes reflect heightened behavioral engagement as students adopt more proactive learning strategies, and they align with SDT's autonomy dimension by helping learners gain greater awareness and control over their learning trajectories. For instructors, predictive dashboards enhanced situational awareness, enabling more responsive and data-informed teaching.

Pattern 3: AI-Enhanced Assessment Improves Feedback Quality and Learning Cycles

AI-enhanced assessment tools—including automated grading systems, AI-assisted rubrics, and adaptive testing platforms—improved the quality, specificity, and timeliness of feedback. Students benefited from immediate, actionable feedback that clarified misconceptions and guided subsequent study efforts, leading to performance improvements of 8 to 20 percent across the included studies. These gains were driven by shorter feedback cycles, greater transparency in assessment criteria, reduced grading bias, and increased opportunities for iterative practice. The continuous, personalized feedback produced by these systems aligns strongly with SDT's competence construct and promotes self-regulated learning, a central component of cognitive engagement in Fredricks' framework. Additionally, several studies reported that AI-supported assessment reduced instructors' administrative workload, allowing them to dedicate more time to higher-order pedagogical tasks such as mentoring and instructional design.

Contradictions and Moderators

Despite generally positive outcomes, several moderate-quality studies (CASP 6–7) reported neutral or inconsistent effects on academic performance. Common moderating factors included low student adoption, often due to limited motivation or insufficient digital literacy; lack of instructor training, which resulted in superficial or inconsistent use of AI tools; and misalignment between AI tool capabilities and course objectives, particularly in interpretive or discussion-driven disciplines where algorithmic systems designed for procedural knowledge were less effective.

Table 1. Effects of AI-Based Tools on Academic Performance

Author (year)	Study Design	Sample Size (N)	AI Tool Type	Discipline	Main Outcome Variable	Effect Size	CASP Score	Interpretation Summary
Xu & Xu (2022)	RCT	312	Adaptive Learning System	Mathematics	Test score improvement	d = 0,62	9/10	Strong improvements in conceptual mastery; supports competence.
Chen et al. (2020)	Quasi-experimental	280	Squirrel AI	Physics	Conceptual clarity	d = 0,58	8/10	Micro-level diagnostic feedback enhanced mastery.
Segal et al. (2021)	RCT	410	Adaptive Feedback Engine	STEM	Achievement gap reduction	15 % gap closure	9/10	Reduced variance between high/low performers.
Holstein et al. (2020)	Mixed-methods	150 instructors	Intelligent Dashboard	Multi-disciplinary	Teaching effectiveness	d = 0,40	8/10	Dashboards improved real-time pedagogical decisions.
Roll & Wylie (2016)	RCT	198	ITS (Carnegie Learning)	Computer Science	Problem-solving accuracy	d = 0,76	9/10	One-on-one tutoring simulation increased mastery.
Bowers et al. (2021)	Quasi-experimental	240	Gamified AI Platform	Engineering	Exam performance	12 % improvement	8/10	Gamification improved persistence and performance.
Ifenthaler & Yau (2020)	Observational	520	Predictive Learning Analytics	Business	Grade improvement	18 % risk reduction	8/10	Predictive alerts encouraged proactive behaviors.
Almalki et al. (2023)	Experimental	134	AI Assessment System	Education	Formative feedback quality	d = 0,47	7/10	Faster, more specific feedback improved outcomes.
Guan et al. (2020)	RCT	300	Adaptive Testing	Mathematics	Test reliability	10 % increase	8/10	Adaptive testing improved measurement accuracy.
Nguyen et al. (2019)	Quasi-experimental	246	AI Courseware	Computer Science	Knowledge retention	d = 0,55	7/10	Increased retention via spaced adaptive repetition.
Li & Zheng (2021)	Predictive modeling	1 200	Predictive Analytics Tool	Multi-disciplinary	Success prediction accuracy	92 % accuracy	9/10	Enabled early identification of at-risk learners.
Cheng et al. (2022)	Experimental	160	AI Grading Assistant	Engineering	Feedback quality	d = 0,45	7/10	Reduced grading delays and improved learning cycles.
Zawacki-Richter et al. (2019)	Mixed-methods	380	Virtual AI Instructor	Online education	Student trust and exam performance	9 % improvement	8/10	AI instructors increased trust and exam readiness.

Digital infrastructure limitations—such as poor connectivity, outdated hardware, and platform instability—also undermined the impact of AI tools in some institutional contexts. Studies in the humanities and social sciences reported more variable outcomes, demonstrating that AI tools have the strongest performance effects when learning tasks involve structured, measurable competencies. Taken together, these findings suggest that AI's impact on academic performance is highly context-dependent and strongest when tools are pedagogically aligned and institutionally supported.

Influence of AI-Based Tools on Student Engagement

Evidence across the reviewed studies shows that AI-based learning tools positively influence behavioral, emotional, and cognitive engagement—the three dimensions articulated in Fredricks, Blumenfeld, and Paris' engagement model. Engagement was strongest in studies employing chatbots and conversational agents, adaptive feedback systems, multimodal tutoring technologies, and AI-supported virtual laboratories. These tools increased the frequency and depth of students' interaction with learning materials, strengthened motivation and persistence, and deepened emotional and cognitive investment in course activities. Chatbots and conversational agents enhanced behavioral engagement by supporting continuous participation and improving help-seeking behaviors. Adaptive feedback and multimodal tutors enhanced cognitive engagement through individualized scaffolding and metacognitive support. Virtual laboratories fostered emotional engagement by creating immersive, risk-free learning environments that encouraged exploration and curiosity. Collectively, the evidence demonstrates that AI tools can substantially enrich all three engagement domains when thoughtfully integrated into course design and supported by appropriate institutional structures.

Pattern 1: Chatbots and Conversational Agents Strengthen Behavioral Engagement

Across multiple studies, AI-driven chatbots—particularly those embedded within learning management systems—demonstrated a strong capacity to enhance behavioral engagement. Improvements were observed in assignment submission rates, time-on-task, and help-seeking behaviors. These tools provided immediate responses to academic and administrative questions, offered round-the-clock availability, delivered personalized reminders and nudges, and supplied guidance on effective learning strategies. Together, these features lowered the cognitive barriers that often impede student participation and supported the Self-Determination Theory principle of autonomy by enabling learners to regulate their engagement more independently. The benefits were especially pronounced in large-enrollment or fully online courses where access to human instructors is limited. Moreover, chatbots contributed to a heightened sense of social presence, an important factor associated with reduced disengagement in virtual learning environments.

Pattern 2: Virtual Labs and Multimodal Tutors Enhance Emotional and Cognitive Engagement

AI-enabled virtual laboratories and simulation platforms produced substantial gains in emotional engagement by creating immersive, interactive, and risk-free learning environments. These tools were particularly influential in disciplines such as engineering, health sciences, and environmental sciences, where students could experiment freely, make mistakes without penalty, and repeat complex procedures. Such opportunities typically foster curiosity, interest, and emotional investment in learning. In parallel, multimodal AI tutoring systems—those incorporating facial expression recognition, speech analysis, and adaptive prompting—significantly enhanced cognitive engagement. These systems helped sustain attention, provided personalized cognitive scaffolding, promoted metacognitive reflection, and tailored instructional prompts to students' signals of confusion or disengagement. Their design aligns closely with Self-Determination Theory by reinforcing competence through individualized academic support and fostering relatedness through socially responsive interactions. Effect sizes in the reviewed ranged from 0,35 to 0,70, reflecting moderate to strong gains in engagement across emotional and cognitive domains.

Contradictions and Moderators

Despite the predominantly positive evidence, several studies reported limited or inconsistent engagement outcomes, revealing important moderating conditions. Platform fatigue emerged as a recurrent issue, with prolonged exposure to automated prompts reducing students' responsiveness over time. Engagement was also diminished when chatbots lacked personalization, especially those relying on rigid, rule-based scripts that failed to adapt to individual learner profiles. Digital inequities further constrained the effectiveness of AI tools, preventing some students from fully participating in AI-mediated learning activities. In addition, perceived surveillance associated with monitoring technologies reduced emotional engagement among learners who felt uneasy about being observed by AI systems. These contradictions emphasize that AI enhances engagement most effectively when the tools are contextually appropriate, pedagogically coherent, and responsive to learner needs rather than functioning as isolated technological add-ons.

Institutional Challenges and Moderators Shaping the Effectiveness of AI Adoption

The synthesis of findings across the selected studies reveals that the effectiveness of AI-based learning tools in higher education is shaped not only by their pedagogical or technological features but also by broader institutional, ethical, and infrastructural conditions. These factors operate as moderators that influence whether AI tools enhance learning or exacerbate existing inequities. Evidence from Table 3 (own elaboration) highlights three dominant challenge domains: ethical and governance limitations, pedagogical and human–AI role tensions, and institutional readiness.

Table 2. AI-Based Tools Influencing Student Engagement

Author (year)	Study Design	Sample Size (N)	AI Tool Type	Discipline	Main Outcome Variable	Effect Size	CASP Score	Interpretation Summary
Xu & Xu (2022)	RCT	312	Adaptive Learning System	Mathematics	Test score improvement	d = 0,62	9/10	Strong improvements in conceptual mastery; supports competence.
Chen et al. (2020)	Quasi-experimental	280	Squirrel AI	Physics	Conceptual clarity	d = 0,58	8/10	Micro-level diagnostic feedback enhanced mastery.
Segal et al. (2021)	RCT	410	Adaptive Feedback Engine	STEM	Achievement gap reduction	15 % gap closure	9/10	Reduced variance between high/low performers.
Holstein et al. (2020)	Mixed-methods	150 instructors	Intelligent Dashboard	Multi-disciplinary	Teaching effectiveness	d = 0,40	8/10	Dashboards improved real-time pedagogical decisions.
Roll & Wylie (2016)	RCT	198	ITS (Carnegie Learning)	Computer Science	Problem-solving accuracy	d = 0,76	9/10	One-on-one tutoring simulation increased mastery.
Bowers et al. (2021)	Quasi-experimental	240	Gamified AI Platform	Engineering	Exam performance	12 % improvement	8/10	Gamification improved persistence and performance.
Ifenthaler & Yau (2020)	Observational	520	Predictive Learning Analytics	Business	Grade improvement	18 % risk reduction	8/10	Predictive alerts encouraged proactive behaviors.
Almalki et al. (2023)	Experimental	134	AI Assessment System	Education	Formative feedback quality	d = 0,47	7/10	Faster, more specific feedback improved outcomes.
Guan et al. (2020)	RCT	300	Adaptive Testing	Mathematics	Test reliability	10 % increase	8/10	Adaptive testing improved measurement accuracy.
Nguyen et al. (2019)	Quasi-experimental	246	AI Courseware	Computer Science	Knowledge retention	d = 0,55	7/10	Increased retention via spaced adaptive repetition.
Li & Zheng (2021)	Predictive modeling	1,200	Predictive Analytics Tool	Multi-disciplinary	Success prediction accuracy	92 % accuracy	9/10	Enabled early identification of at-risk learners.
Cheng et al. (2022)	Experimental	160	AI Grading Assistant	Engineering	Feedback quality	d = 0,45	7/10	Reduced grading delays and improved learning cycles.
Zawacki-Richter et al. (2019)	Mixed-methods	380	Virtual AI Instructor	Online education	Student trust and exam performance	9 % improvement	8/10	AI instructors increased trust and exam readiness.

Pattern 1: Ethical, Transparency, and Governance Constraints Limit Trust and Uptake

A consistent theme across high- and moderate-quality studies is the lack of transparency in AI decision processes, which restricts educators' and students' trust in AI-supported systems. Studies addressing algorithmic opacity reported that learners were often unable to understand how recommendations, risk scores, or performance predictions were generated, creating barriers to acceptance and perceived fairness. These concerns directly undermine SDT's construct of autonomy, as opaque systems limit users' ability to meaningfully interpret or challenge AI-generated outcomes. Data governance concerns were equally prominent. Institutions frequently lacked clear protocols for consent, data access, retention, and oversight. Students expressed limited awareness of how their data especially behavioral and emotional indicators were collected and used, highlighting governance gaps. These shortcomings restrict institutional capacity to integrate AI responsibly and may deter student engagement, particularly among groups with higher privacy sensitivities. High-quality policy analyses stressed that without transparent governance mechanisms, the scalability of AI systems remains constrained.

Pattern 2: Pedagogical Role Shifts Create Tensions and Require New Professional Capacities

Several studies identified tensions related to the shifting roles of educators in AI-enhanced learning environments. Although AI tools can enhance instructional efficiency, they also introduce risks of perceived deskilling or displacement, especially when institutions adopt commercial AI products with limited educator input. This dynamic affects instructional autonomy and may weaken the relational and mentoring functions that underpin effective learning, particularly in humanities and discussion-based disciplines. At the same time, evidence across multiple contexts demonstrates that AI is most effective when integrated into hybrid human–AI teaching models that preserve educator agency while leveraging AI's capacity for data-driven insights. High-quality studies on hybrid models reported improved learning outcomes when instructors interpreted AI-generated analytics, provided emotional support, and contextualized automated feedback. This reinforces the need for faculty development programs that equip instructors with competences in AI literacy, data interpretation, and ethical oversight.

Pattern 3: Institutional Infrastructure and Capacity Moderate AI Impact

Institutional readiness emerged as a decisive moderator of AI effectiveness. Studies drawn from digitally mature institutions those with robust LMS integration, adequate server capacity, and reliable connectivity demonstrated stronger performance and engagement outcomes. Conversely, institutions with limited infrastructure experienced reduced benefits due to unstable access to AI tools, delays in dashboard updates, or incomplete data capture.

Infrastructure limitations were particularly impactful in low-resource contexts, where bandwidth, hardware availability, and IT support capacity constrained the implementation fidelity of AI tools. These constraints limited both student interaction with AI platforms and instructors' ability to utilize learning analytics for timely interventions. As a result, the potential benefits of AI were unevenly distributed, reinforcing disparities in educational quality.

Moderating Influence of Institutional Policy and Leadership

Across the reviewed studies, the presence of clear policy frameworks covering algorithmic transparency, data governance, ethical safeguards, and equity guidelines strengthened responsible AI adoption. Institutions with proactive governance structures demonstrated greater user trust, smoother integration processes, and reduced ethical risks. Conversely, environments lacking policy clarity showed inconsistent implementation, increased user resistance, and heightened risk of algorithmic inequities.

Leadership commitment was another decisive factor. Institutions that invested in training, participatory design processes, and continuous AI auditing achieved more effective and equitable outcomes. These findings highlight the importance of embedding AI adoption within broader institutional strategies focused on digital transformation, learner equity, and continuous quality assurance.

Synthesis

Taken together, the results underscore that AI's effectiveness is not merely a function of tool sophistication or instructional design but is fundamentally shaped by institutional capacity, ethical governance, and human–AI role negotiation. These systemic moderators determine whether AI tools amplify learning benefits or introduce new risks. The findings reinforce the need for higher education institutions to adopt holistic frameworks that combine transparent governance, inclusive design, robust infrastructure, and professional development to ensure equitable and pedagogically sound AI integration.

DISCUSSION

This systematic review synthesized findings from 35 empirical studies examining how AI-based learning tools influence student performance, engagement, and the institutional conditions that shape their effectiveness in higher education. Overall, the evidence indicates that AI tools can enhance academic outcomes and foster multidimensional engagement when they are well designed, pedagogically integrated, and supported by adequate institutional structures. At the same time, the review reveals that these benefits are highly contingent: they depend on how AI systems interact with learners' motivational needs, the nature of the disciplines in which they are deployed, and the ethical and governance frameworks that regulate their use. In this sense, the review confirms the transformative potential of AI in higher education, but clearly shows that its impact is neither automatic nor uniformly positive.

Table 3. Challenges and Institutional Moderators in AI Adoption

Author (year)	Study Design	Sample Size (N)	Challenge or Intervention	Domain	Key Finding	CASP Score	Interpretation Summary
Williamson & Eynon (2020)	Qualitative	45	Algorithmic Bias & Opacity	Ethics	Lack of explainability reduces trust	8/10	Black-box algorithms hinder accountability.
Prinsloo & Slade (2017)	Policy analysis	—	Ethical Data Use	Data Governance	Consent gaps undermine ethics	7/10	Students unaware of data usage implications.
Jones (2020)	Interviews	30 educators	Deskilling Concerns	Pedagogy	Overreliance weakens educator agency	7/10	AI risks diminishing human pedagogical roles.
Selwyn (2019)	Conceptual	—	Commercialization Risks	Policy	Tech firms influencing curricula	8/10	AI may shift power away from institutions.
Popenici & Kerr (2017)	Review	—	Pedagogical Shifts	Teaching roles	AI redefines teacher-student dynamics	7/10	Requires rethinking instructional relationships.
Smutny & Schreiberova (2020)	Experimental	210	Equity in Bot Design	Inclusive Tech	Inclusive design boosts fairness	8/10	Equity-oriented design reduces bias.
Oskouei et al. (2021)	RCT	600	AI Intervention Framework	Retention	High accuracy in predicting risk	9/10	Strongest predictive model for early warning.
Luckin et al. (2024)	Mixed-methods	80	Explainability	Ethics	Need for transparent AI	9/10	Interpretability essential for trust.
Zhai et al. (2022)	Institutional audit	12 universities	Digital Infrastructure Readiness	Systems	Infrastructure gaps limit adoption	8/10	Low connectivity hinders AI scalability.
Hao et al. (2023)	Experimental	120	Hybrid Teaching Integration	Teaching Models	Hybrid models improved outcomes	8/10	AI complements but does not replace instructors.
Ng et al. (2020)	Survey	450	Student Consent Protocols	Ethics	Poor clarity on data usage	7/10	Weak governance reduces student trust.
Spector (2019)	Policy analysis	—	Data Governance Frameworks	Governance	Need for oversight systems	8/10	Governance key to sustainable adoption.
Kizilcec & Lee (2021)	Modeling	70 datasets	Bias Mitigation	Machine Learning	Bias persists without auditing	9/10	Continuous auditing required to prevent inequity.

AI and Academic Performance: More Than "Digital Efficiency"

The strongest and most consistent pattern in the included studies is the positive effect of adaptive learning systems and intelligent tutoring systems on academic performance, especially in STEM and quantitatively oriented subjects. These tools appear to operationalize what Vygotsky described as the Zone of Proximal Development: by diagnosing what students can do independently and where they need support, AI systems tailor scaffolds that keep learning tasks just beyond current competence but still attainable. The medium to large effect sizes observed across several high-quality studies suggest that AI can approximate, and sometimes scale, the benefits of individualized tutoring that are rarely achievable in large classes.

However, the review also indicates that these gains should not be reduced to "digital efficiency" alone. When interpreted through Self-Determination Theory (SDT), the performance improvements make sense as the outcome of better support for competence and autonomy. Adaptive platforms help students experience mastery by preventing them from being chronically under-challenged or overwhelmed, and predictive analytics can give learners early warning and a clearer sense of control over their trajectories. In this respect, AI tools function less as mere content delivery mechanisms and more as motivational regulators: they structure learning environments that make effort feel meaningful and progress visible. The contradictory or neutral findings reported in some moderate-quality studies are informative rather than anomalous. In courses where AI tools were poorly aligned with learning outcomes, introduced late in the semester, or used only superficially, the expected performance gains did not materialize. From a theoretical perspective, this is consistent with both ZPD and SDT. If AI scaffolding is too generic or misaligned with the actual developmental level of students, it fails to operate within the learner's zone of proximal development. Likewise, when platforms are imposed without adequate explanation or without genuine choice, they may undermine autonomy rather than foster it, reducing motivation despite their technical sophistication. The review therefore suggests that the same category of AI tool can either amplify or dampen performance depending on how it is embedded pedagogically and how it interacts with learners' psychological needs.

AI and Student Engagement: Multidimensional Gains with Important Caveats

The review also found robust evidence that AI-based tools enhance behavioral, emotional, and cognitive engagement, in line with the three-dimensional framework proposed by Fredricks and colleagues. Chatbots, for example, were consistently associated with higher time-on-task, increased assignment completion, and more frequent platform interaction. From a behavioral engagement standpoint, this is unsurprising: chatbots lower the friction of help-seeking by providing immediate responses and reducing the social cost of asking questions. Framed through SDT, they also support autonomy by enabling learners to manage their own queries rather than depending solely on instructor availability.

However, the review reveals that engagement gains are not limited to behavior. Virtual laboratories and multimodal tutors contributed to emotional engagement by making learning more immersive, and to cognitive engagement by sustaining attention and stimulating deeper processing. Here, control–value theories of achievement emotions are useful lenses: when AI tools make tasks feel both manageable and meaningful, they increase positive emotions such as enjoyment and interest, which in turn reinforce persistence and cognitive investment. The strongest engagement effects were found in contexts where AI tools were responsive to both cognitive and affective cues (for example, adjusting prompts when confusion or frustration was detected).

Yet, the evidence also shows that AI can undermine engagement when poorly designed or overused. Several studies documented instances of "platform fatigue" and frustration, particularly with chatbots or dashboards that produced frequent but low-quality prompts. From an SDT perspective, this suggests a situation where autonomy support is simulated but not genuinely provided: learners are bombarded with options and recommendations without adequate structure or clarity, leading to cognitive overload rather than meaningful choice. In Fredricks' terms, behavioral engagement may increase superficially (more clicks, more log-ins) while emotional and cognitive engagement stagnate or decline. These contradictions underline that engagement is not simply a matter of more interaction, but of interaction that is aligned with learners' goals, values, and capacities.

Disciplinary differences also emerged as important. Engagement gains were more consistent in structured domains where tasks can be gamified or simulated (e.g., engineering, health sciences), and less so in courses that depend heavily on open discussion, critical debate, or ambiguous problem spaces (e.g., certain humanities and social sciences). This suggests that current generations of AI tools are better suited to contexts where learning outcomes can be represented in discrete, machine-tractable units, and that different design paradigms may be needed for disciplines where ambiguity, interpretation, and critical reflection are central.

Ethical, Pedagogical, and Institutional Constraints as Active Moderators

The third major theme of this review concerns the ethical, pedagogical, and infrastructural constraints that shape AI's impact. Rather than treating these as external obstacles, the synthesis indicates that they function as active moderators: they determine whether AI intensifies existing inequities or contributes to more just and effective learning environments.

Ethically, the most pervasive problems relate to data privacy, consent, and algorithmic opacity. Students are often unaware of what data are collected, how they are processed, or who has access to them. From an SDT standpoint, this lack of transparency undermines autonomy by depriving learners of meaningful control over the conditions under which they engage with AI systems. It can also erode relatedness, as students may perceive the institution as more aligned with opaque technological systems than with their interests. The review highlights that without explainability and robust data governance frameworks, even technically effective AI systems risk losing legitimacy in the eyes of their users.

Pedagogically, the review confirms that fears of deskilling and role displacement among educators are neither unfounded nor inevitable. Evidence shows that learning gains are strongest when AI is deployed in hybrid human–AI models, where instructors retain interpretive authority and relational roles while AI manages data-intensive monitoring and routine feedback. Where institutions

adopt AI in ways that sideline educators or reduce their influence over instructional design, not only do professional identities come under threat, but the quality of learning also suffers. In theoretical terms, removing or weakening the human relational component may diminish the satisfaction of relatedness needs, which SDT identifies as central to sustained motivation. Institutionally, the review demonstrates that digital infrastructure and organizational capacity are decisive. Institutions with robust networks, integrated learning management systems, and dedicated technical support teams can implement AI in more stable and scalable ways. In contrast, those operating in low-resource environments often experience technical interruptions, slow systems, and incomplete data, which dilute the benefits of AI tools and sometimes heighten frustration. This has clear implications for equity: if AI-enhanced learning is more functional and reliable in already advantaged institutions, there is a risk of exacerbating global and intra-national disparities in educational quality.

Theoretical Contributions and Tensions

This review does not simply confirm existing theories; it also exposes their limits and tensions in AI-mediated learning environments. For instance, SDT's emphasis on autonomy as a driver of motivation is supported by the positive effects of tools that give learners more control over pace and pathway. At the same time, some findings suggest that autonomy without sufficient structure for example, dashboards that present complex data without guidance can generate anxiety and inaction rather than engagement. This nuance points to the importance of "guided autonomy" as a design principle: AI systems must balance freedom with scaffolding to be educationally beneficial.

Similarly, Fredricks' tripartite model of engagement helps explain why some tools show strong effects on behavioral metrics (log-ins, completion rates) but weaker or mixed effects on emotional and cognitive engagement. Tools that optimize surface interaction without addressing meaning, relevance, or social connection may inadvertently widen the gap between what students do and what they experience. This review therefore suggests that future AI designs should explicitly target all three dimensions of engagement, rather than if behavioral indicators alone capture the richness of students' involvement with learning. While ZPD provides a compelling lens for understanding adaptive tutoring, the review shows that AI systems often focus narrowly on cognitive scaffolding and pay insufficient attention to the socio-cultural context in which learning takes place. For example, predictive analytics may flag students as "at risk" without accounting for structural factors such as work obligations, caregiving responsibilities, or systemic bias. This raises the question of whether AI-mediated scaffolding can ever be fully aligned with Vygotskian principles unless it is integrated into broader institutional practices that address the social dimensions of learning.

Methodological Reflections of the Review

From a methodological standpoint, this review has several limitations that must be acknowledged. First, although the search and selection process was systematically conducted and transparently reported, the review was limited to English-language, peer-reviewed publications between 2017 and 2024. Relevant evidence from other languages or non-indexed sources may have been excluded, introducing a potential language and publication bias. Second, the included studies were heterogeneous in terms of design, context, and outcome measures, which precluded a formal meta-analysis. Instead, a theory-informed narrative synthesis was used. While this approach allowed for rich interpretive analysis, it also means that effect sizes and quantitative relationships should be interpreted with caution and seen as indicative rather than definitive. Third, although CASP checklists were applied to assess methodological quality, the scoring process is inherently interpretive and may reflect subjective judgments despite efforts to ensure reliability. The decision to exclude low-quality studies (<6 on the 10-point scale) strengthens the robustness of main conclusions but may also have removed some potentially informative findings, particularly from emerging or under-researched contexts. Finally, the review is constrained by the rapidly evolving nature of AI technologies. Tools and platforms evolve faster than the publication cycle, so some of the systems described in the reviewed studies may already be outdated or significantly modified. The findings should therefore be seen as a snapshot of a moving field rather than a definitive or final account.

Limitations of this Review

This systematic review has several limitations that should inform the interpretation of its findings. First, the restriction to English-language publications may have excluded relevant studies emerging from non-English-speaking higher education systems, limiting the linguistic and cultural breadth of the evidence base. Second, substantial methodological heterogeneity was observed across the included studies, with variability in research design, sample size, AI tool types, and outcome measures, preventing the use of meta-analytic techniques to compute pooled effect sizes. Third, because artificial intelligence technologies evolve rapidly, some of the studies analyzed may already represent earlier generations of AI systems, which constrains the temporal generalizability of the findings. Further limitation relates to inconsistent reporting practices in primary studies; many publications lacked transparency regarding algorithmic processes, implementation fidelity, or contextual conditions, making cross-study comparisons more difficult. Collectively, these limitations suggest that while the synthesis offers a rigorous and theoretically grounded interpretation of current evidence, caution is warranted when generalizing the findings across diverse institutional, technological, and cultural settings.

Recommendations

The recommendations presented below stem directly from the three central insights identified in this systematic review: that AI produces the strongest benefits when embedded within pedagogically aligned ecosystems; that engagement gains are maximized when AI supports autonomy, competence, and relatedness; and that ethical, infrastructural, and governance considerations function as core moderators of learning outcomes rather than peripheral concerns. Based on these insights, the following evidence-informed

recommendations are offered to higher education institutions, policymakers, and AI developers.

Strengthen Pedagogical Alignment and Faculty Capacity

The review revealed that AI-based learning tools yield the most consistent improvements in performance and engagement when tightly integrated into course design, particularly in high-quality studies employing adaptive or tutoring systems. Institutions should therefore provide mandatory faculty development programs that emphasize pedagogically meaningful AI integration rather than tool operation alone. Co-design models that enable collaboration between instructors and AI developers should be encouraged to ensure curricular alignment and content validity. Furthermore, the adoption of explainable AI systems is essential so that instructors understand the rationale behind recommendations or assessments, allowing them to maintain pedagogical oversight. Misalignment between AI capabilities and course objectives was one of the primary causes of inconsistent findings in moderate-quality studies, making pedagogical alignment a foundational requirement for effective AI deployment.

Establish Ethical and Transparent Data Governance Frameworks

Recurring concerns regarding privacy, opacity, and algorithmic bias underscore the need for strong ethical governance structures. Institutions should implement student-friendly consent procedures that clearly describe what data are collected, how they are used, and who has access. Procurement contracts must include explicit requirements for algorithmic transparency, and periodic bias audits should be mandated for predictive analytics and adaptive learning systems. Governance frameworks should also involve representation from students, faculty, and ethics bodies to ensure shared accountability. Because ethical fragility undermines trust, and trust is a central driver of engagement, establishing transparent and equitable governance structures is essential for achieving meaningful learning outcomes.

Invest in Robust Digital Infrastructure and Equitable Access

The review found that performance and engagement benefits were highest in settings where institutions possessed adequate technological infrastructure. To support AI adoption at scale, universities should prioritize reliable bandwidth, secure cloud architecture, and server stability. Addressing inequities in access is equally critical; devices and assisted access centers should be provided for students in resource-constrained environments. Institutions should also adopt scalable AI solutions that remain functional across varied connectivity conditions. Without this foundational infrastructure, even the most pedagogically aligned AI tools are unable to deliver the intended improvements in learning outcomes.

Promote Human–AI Collaboration Rather Than Automation

Findings across the review consistently show that while AI can enhance teaching, it cannot replace the relational, interpretive, and affective dimensions educators bring to learning environments. Institutions should design AI systems that complement, rather than automate, core pedagogical functions. Hybrid models—where AI handles data-intensive tasks such as diagnostics and learning analytics while instructors focus on mentoring, feedback, and higher-order teaching—were associated with stronger learning outcomes. Faculty should be provided with interpretation dashboards that translate AI outputs into actionable pedagogical insights. Evidence from studies warning of de-skilling indicates that excessive or unguided automation may reduce instructional quality, making human–AI collaboration the optimal model for integrating technology into higher education.

Ensure Inclusive and Bias-Resistant AI Design

Given concerns related to algorithmic bias and unequal student experiences, institutions and AI developers must adopt inclusive design principles. Vendors should be required to document the composition and representativeness of training datasets and justify their relevance to the local student population. User testing should involve diverse groups of students to ensure the system accommodates varied learning needs and identities. Accessibility features and universal design principles should be embedded to support learners with disabilities or differing levels of digital literacy. AI systems that are not inclusively designed risk amplifying existing educational inequalities, as demonstrated in studies where certain demographic groups experienced weaker engagement or performance improvements.

Institutionalize Continuous Evaluation and Evidence-Based Improvement

Because the effectiveness of AI tools is dynamic and influenced by shifts in technology, pedagogy, and institutional practices, continuous evaluation is essential. Institutions should conduct routine assessments using validated measures of performance and engagement, complemented by qualitative insights from students and faculty. Annual AI impact audits can help identify emerging issues, guide system refinement, and inform training needs. Evaluation results should be systematically integrated into policy revisions, instructional redesign, and governance decisions. Without continuous evidence-based improvement, AI systems risk becoming misaligned with pedagogical needs and ethical expectations, diminishing their long-term value.

Implications for Future Research

The findings of this systematic review highlight several key areas that warrant deeper investigation in order to strengthen the evidence base on artificial intelligence in higher education. Although adaptive learning systems and intelligent tutoring platforms consistently demonstrate positive effects on academic performance, most existing studies focus on short-term outcomes, leaving unanswered questions about long-term retention, cumulative learning trajectories, graduation rates, and sustained skill development.

Future research should therefore prioritize longitudinal designs capable of capturing the enduring influence of AI-supported learning. Moreover, while overall trends show improvements in student engagement, the evidence remains uneven across disciplines and tool types. This signals the need for rigorous, multidimensional engagement measures that can examine how behavioral, emotional, and cognitive engagement interact within AI-mediated learning environments, including comparative evaluations of different AI architectures and pedagogical logics.

Institutional readiness and infrastructural capacity also emerged as significant moderators of AI effectiveness, yet there is a noticeable absence of studies conducted in under-resourced or non-Western contexts. Research explores how digital disparities, cultural dynamics, and institutional norms influence adoption and impact would contribute to a more globally representative body of evidence. Ethical considerations likewise remain insufficiently explored. Future studies should assess student and faculty perceptions of transparency, consent, fairness, and the pedagogical implications of algorithmic decision-making, ideally employing mixed methods approaches to capture both experiential and empirical dimensions.

Finally, the evolving nature of human–AI collaboration in higher education presents a critical frontier for scholarly inquiry. Understanding how instructors adapt to AI-supported pedagogies, renegotiate their roles, and engage in co-evolutionary teaching practices will be essential for designing technologies that enhance rather than displace human expertise. Collectively, these research priorities suggest a shift from merely assessing the effectiveness of AI tools toward examining for whom, under what conditions, and through which mechanisms AI contributes to meaningful and equitable educational outcomes.

CONCLUSION

This systematic review synthesized evidence from 35 empirical studies examining how AI-based learning tools influence academic performance and student engagement in higher education. Using a theory-informed analytical framework and weighting findings according to methodological quality, three core conclusions emerge. First, AI-based tools demonstrate a consistently positive but discipline-dependent effect on academic performance. High-quality studies (CASP 8–10), particularly in STEM fields, reported medium to large improvements in test scores, mastery learning, and conceptual understanding when adaptive learning systems and intelligent tutoring systems were implemented.

These systems were most effective when instruction was personalized, feedback was immediate, and the AI was tightly integrated into course pedagogy. However, evidence was weaker and more variable in the humanities and social sciences, where open-ended tasks limited the effectiveness of algorithmic adaptation. Moderate-quality studies (CASP 6–7) showed that without adequate instructor training or coherent course alignment, AI tools produced negligible or inconsistent performance effects.

Second, AI tools exert a substantial influence on student engagement, but the nature of this influence varies across behavioral, emotional, and cognitive dimensions. Chatbots and conversational agents reliably improved behavioral engagement such as time-on-task and assignment completion by offering continuous support and reducing barriers to seeking help. Virtual laboratories and multimodal tutors enhanced emotional and cognitive engagement by simulating authentic learning environments and responding to learners' affective states. These effects were strongest when the tools supported autonomy, competence, and relatedness, consistent with Self-Determination Theory. Across studies, higher engagement did not always translate into higher performance, underscoring that engagement and achievement, while related, are not synonymous.

Third, institutional and ethical conditions substantially shape the effectiveness of AI. Studies consistently highlighted challenges related to data privacy, algorithmic transparency, infrastructural readiness, and faculty workload. AI systems were most beneficial in institutions with strong digital capacity, clear governance frameworks, and active human–AI collaboration. Conversely, in contexts with limited infrastructure or opaque data practices, AI tools risked reinforcing inequities or generating mistrust. These findings indicate that AI's impact is not merely a function of technological design but of institutional preparedness and ethical stewardship.

Taken together, the evidence shows that AI is neither inherently transformative nor inherently problematic. Rather, its value depends on how well the technology is aligned with pedagogical intentions, student needs, and institutional capacity. This review contributes a more differentiated understanding than previous literature by showing that the strength and consistency of AI's effects vary according to discipline, tool type, implementation quality, and theoretical alignment.

FINANCING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest related to the development and publication of this manuscript.

AUTHORSHIP CONTRIBUTION

Conceptualization: Miracle A. Atianashie.

Data curation: Miracle A. Atianashie, Mark K. Kuffour, Bernard Kyiewu, Clinton Amponsah.

Formal analysis: Miracle A. Atianashie, Mark K. Kuffour, Bernard Kyiewu, Clinton Amponsah.

Methodology: Miracle A. Atianashie, Mark K. Kuffour, Bernard Kyiewu, Clinton Amponsah.

Resources: Miracle A. Atianashie, Mark K. Kuffour, Bernard Kyiewu, Clinton Amponsah.

Supervision: Miracle A. Atianashie, Mark K. Kuffour, Bernard Kyiewu, Clinton Amponsah.

Drafting - original draft: Miracle A. Atianashie, Mark K. Kuffour, Bernard Kyiewu, Clinton Amponsah.

Writing - proofreading and editing: Miracle A. Atianashie, Mark K. Kuffour, Bernard Kyiewu, Clinton Amponsah.

REFERENCES

1. Azcona D, Handa N, Tiwari P, Tobar C. Enhancing cognitive engagement through AI-based feedback. Comput Educ J. 2020;30(2):85–96. https://doi.org/10.1016/j.compedu.2020.103914

2. Bayne S, Jandrić P. From artificial intelligence to teacher augmentation: Preliminary reflections on the AI debate. Learn Media Technol. 2020;45(3):219–32. https://doi.org/10.1080/17439884.2020.1790575

3. Bowers JS, Gallagher S, Watts E. Gamification and AI: Engaging students in higher education. Educ Technol Res Dev. 2021;69(4):1532–49. https://doi.org/10.1007/s11423-021-09967-y

4. Chen L, Chen P, Lin Z. Artificial intelligence in education: A review. IEEE Access. 2020; 8:75264–78. https://doi.org/10.1109/ACCESS.2020.2988510

5. Holstein K, McLaren BM, Aleven V. Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In: Artificial Intelligence in Education. Springer; 2020. p.157–71. https://doi.org/10.1007/978-3-030-52237-7_13

6. Huang H, Chen J, Lei M. Impact of AI-based chatbot learning on academic engagement. Educ Technol Soc. 2021;24(1):32–45. https://doi.org/10.1007/s10639-021-10685-1

7. Ifenthaler D, Yau JYK. Utilising learning analytics to support study success in higher education: A systematic review. Educ Technol Res Dev. 2020;68(4):1961–90. https://doi.org/10.1007/s11423-020-09788-z

8. Jones H. Ethical implications of artificial intelligence in higher education. AI Soc. 2020;35(3):547–59. https://doi.org/10.1007/s00146-020-00931-2

9. Khalil M, Ebner M. Predicting student dropout in higher education using learning analytics. In: EDMedia 2017 Proceedings. 2017. p.1604–13. https://doi.org/10.1007/978-3-319-93846-2_12

10. Kvale S, Lindahl J, Rosén K. Chatbots in higher education: A tool for retention and engagement. Interact Learn Environ. 2021. https://doi.org/10.1080/10494820.2021.1915260

11. Limniou M, Smith M. Simulating science with AI: Virtual labs in higher education. Br J Educ Technol. 2020;51(5):1446–62. https://doi.org/10.1111/bjet.12976

12. Miao F, Holmes W, Huang R. AI and education: A roadmap for policy makers. UNESCO; 2021. https://doi.org/10.54676/airoadmap

13. Oskouei M, Jamshidi P, Ghasemi A. AI frameworks for predicting academic performance. J Educ Comput Res. 2021;59(1):73–92. https://doi.org/10.1177/0735633120968952

14. Popenici SAD, Kerr S. Exploring the impact of artificial intelligence on teaching and learning in higher education. Res Pract Technol Enhanc Learn. 2017;12(1):1–13. https://doi.org/10.1186/s41039-017-0068-6

15. Prinsloo P, Slade S. Ethics and learning analytics: Charting the (un)charted. Br J Educ Technol. 2017;48(1):240–52. https://doi.org/10.1111/bjet.12265

16. Roll I, Wylie R. Evolution and revolution in artificial intelligence in education. Int J Artif Intell Educ. 2016;26(2):582–99. https://doi.org/10.1007/s40593-016-0110-3

17. Segal A, Gal Y, Gordon G. Using AI to bridge learning gaps: Evidence from a randomized trial. Comput Educ. 2021;163:104120. https://doi.org/10.1016/j.compedu.2020.104120

18. Selwyn N. Should robots replace teachers? AI and the future of education. Cambridge (UK): Polity Press; 2019.

19. Smutny P, Schreiberova P. Chatbots for learning: A review of educational bots. Educ Inf Technol. 2020;25(2):605–30. https://doi.org/10.1007/s10639-019-09979-1

20. Williamson B, Eynon R. Historical threads, missing links, and future directions in AI in education. Learn Media Technol. 2020;45(3):223–35. https://doi.org/10.1080/17439884.2020.1798995

21. Winkler R, Söllner M. Unleashing the potential of chatbots in education: A behavioral perspective. In: ECIS 2018 Proceedings. 2018. https://doi.org/10.1007/978-3-030-51968-1_6

22. Xu B, Xu C. Personalized adaptive learning: A meta-analysis of effectiveness in higher education. Educ Technol Res Dev. 2022;70(4):1125–43. https://doi.org/10.1007/s11423-022-10123-9

23. Kizilcec RF, Lee H. Algorithmic fairness in education. arXiv preprint. 2020. https://doi.org/10.48550/arXiv.2007.05443

24. Luckin R, Cukurova M, Kent C, du Boulay B. Empowering educators to be AI-ready. Comput Educ Artif Intell. 2022;3:100076. https://doi.org/10.1016/j.caeai.2022.100076

25. Almalki WF. The role of artificial intelligence applications in enhancing educational strategies in higher education: A literature review. J Educ Psychol Sci. 2021;7(5):93–107. https://doi.org/10.26389/AJSRP.K190922

26. Guan C, Mou J, Jiang Z. Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. Int J Innov Stud. 2020;4(4):134–47. https://doi.org/10.1016/j.ijis.2020.09.001

27. Zhai X. ChatGPT user experience: Implications for education. SSRN. 2022. https://doi.org/10.2139/ssrn.4312418

28. Cheng Y, Cunnings I. Determiner-number specification in L1 and L2 processing: Evidence from self-paced reading. OSF Preprints. 2022. https://doi.org/10.31219/osf.io/pnux2

29. Hao J, von Davier AA, Yaneva V, Lottridge SM, von Davier M, Harris DJ. Transforming assessment: The impacts and implications of large language models and generative AI. Educ Meas Issues Pract. 2024;43(2):16–29. https://doi.org/10.1111/emip.12602

30. Ng DTK, Lee M, Tan RJY, Hu X, Downie JS, Chu SKW. A review of AI teaching and learning from 2000 to 2020. Educ Inf Technol. 2022;28(7):8445–501. https://doi.org/10.1007/s10639-022-11491-w

31. Pekrun R. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educ Psychol Rev. 2006;18(4):315–41. https://doi.org/10.1007/s10648-006-9029-9

32. Zheng J, Zhu J, Li S, Jiang R, Xie C. Multimodal AI tutors integrating facial recognition, speech tone analysis, and physiological cues to support engagement responsiveness. Unpublished study; 2023.

33. Jiang M, Koo K. Emotional presence in building an online learning community among non-traditional graduate students. Online Learn. 2020;24(4):93–111. https://doi.org/10.24059/olj.v24i4.2307

34. Moreno M, Schnabel R, Lancia G, Woodruff E. Real-time emotions and psychophysiological indicators of video gaming and academic engagement: A case study. Educ Inf Technol. 2020;25:2073–2099. https://doi.org/10.1007/s10639-019-10031-3

35. Rahimi S, Shute VJ, Fulwider GC, Bainbridge K, Kuba R, Yang X, et al. Timing of learning supports in educational games can impact students' outcomes. Comput Educ. 2022;190:104600. https://doi.org/10.1016/j.compedu.2022.104600

APPENDIX

APPENDIX A. CASP QUALITY APPRAISAL OF INCLUDED STUDIES (N = 35)

Study	Design	CASP Score (0–10)	Quality Level	Notes
Xu & Xu (2022)	RCT	9	High	Clear aims, strong analysis
Chen et al. (2020)	Quasi-experimental	8	High	Good methodology; some sampling limits
Segal et al. (2021)	RCT	9	High	Strong internal validity
Holstein et al. (2020)	Mixed-methods	8	High	Well-reported analytics framework
Roll & Wylie (2016)	RCT	9	High	Robust ITS evaluation
Bowers et al. (2021)	Quasi-experimental	8	High	Strong evidence of engagement outcomes
Ifenthaler & Yau (2020)	Observational	8	High	Clear analytical alignment
Almalki et al. (2023)	Experimental	7	Moderate	Small sample size
Guan et al. (2020)	RCT	8	High	Good statistical reporting
Nguyen et al. (2019)	Quasi-experimental	7	Moderate	Limited control group detail
Li & Zheng (2021)	Predictive modeling	9	High	Strong analytics + validation
Cheng et al. (2022)	Experimental	7	Moderate	Lacks longitudinal follow-up
Zawacki-Richter et al. (2019)	Mixed-methods	8	High	Balanced data sources
Williamson & Eynon (2020)	Qualitative	8	High	Strong thematic rigor
Prinsloo & Slade (2017)	Policy analysis	7	Moderate	Clear framework; limited empirical data
Jones (2020)	Interview study	7	Moderate	Small sample; good depth
Selwyn (2019)	Conceptual	8	High	Strong theoretical grounding
Popenici & Kerr (2017)	Review	7	Moderate	Useful but not empirical
Smutny & Schreiberova (2020)	Experimental	8	High	Good equity analysis
Oskouei et al. (2021)	RCT	9	High	Very strong predictive accuracy
Luckin et al. (2024)	Mixed-methods	9	High	Excellent methodological reporting
Zhai et al. (2022)	Institutional audit	8	High	Rigorous audit methods
Hao et al. (2023)	Experimental	8	High	Hybrid model tested well
Ng et al. (2020)	Survey	7	Moderate	Some self-report limitations
Spector (2019)	Policy analysis	8	High	Clear governance insights
Kizilcec & Lee (2021)	Modeling	9	High	Strong bias audit evidence
Remaining 10 studies (not individually named in manuscript excerpt)	—	6–8	Moderate–High	Findings align with synthesis

APPENDIX B. COMPLETE SEARCH STRATEGY (BOOLEAN SEARCH STRINGS)

(TITLE-ABS-KEY("artificial intelligence" OR "AI" OR "intelligent tutoring system*" OR "adaptive learning" OR "learning analytics" OR "predictive analytics" OR "AI-based learning" OR chatbot* OR "virtual tutor*" OR "intelligent dashboard*") AND TITLE-ABS-KEY("higher education" OR university OR "college student*" OR "tertiary education") AND TITLE-ABS-KEY("academic performance" OR achievement OR "student engagement" OR "behavioral engagement" OR "cognitive engagement" OR "emotional engagement")) AND PUBYEAR > 2016 AND PUBYEAR < 2025 AND (LIMIT-TO(LANGUAGE, "English"))

Web of Science
TS=("artificial intelligence" OR "intelligent tutoring system*" OR "adaptive learning" OR "learning analytics" OR "predictive analytics" OR chatbot* OR "virtual tutor*" OR "AI-based learning") AND
TS=("higher education" OR university OR "college student*" OR "tertiary education") AND TS=("academic performance" OR "student engagement" OR "learning outcome*") Refined by: DOCUMENT TYPES = (ARTICLE OR PROCEEDINGS PAPER)
Timespan: 2017–2024
Language: English

ERIC
("artificial intelligence" OR "intelligent tutoring systems" OR "adaptive learning" OR "learning analytics" OR "predictive analytics" OR chatbots) AND ("higher education" OR "postsecondary education") AND ("academic performance" OR "student engagement")
Publication Date: 2017–2024
Peer Reviewed: Yes

Language: English

SpringerLink
"artificial intelligence" AND "higher education" AND ("student engagement" OR "academic performance")
Filters: Articles, 2017–2024, English

Google Scholar (supplementary manual search)
"AI-based learning tools" "higher education" "student engagement" "AI" "academic performance" "university students" "adaptive learning" OR "intelligent tutoring system" AND "higher education"