

Artificial Intelligence in Physics Education: An Exploratory Systematic Review of Educational Applications, SDG 4 Alignment, and Sustainability Competencies

Chia Shih Su^{1*}, Chuan Chih Hsu², Kua I Su³, Chia Li Su⁴

¹Department of Mathematics, Physics, and Statistics, Universidad Católica del Maule (UCM), Talca, Chile, ²Department of Interdisciplinary Development, General Education Directorate, Universidad Católica del Maule (UCM), Talca, Chile, ³Docotorado en Ciencias Empresariales y Sociales, Universidad de Ciencias Empresariales y Sociales (UCES), Buenos Aires, Argentina, ⁴Hospital General de Agudos Dra. Cecilia Grierson, Buenos Aires, Argentina.

*Corresponding Author: ciaushih@gmail.com

ABSTRACT

This paper presents an exploratory systematic review of artificial intelligence (AI) in physics education and its alignment with Sustainable Development Goal 4 (SDG 4). Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 guidelines (Page et al., 2021), a Web of Science search conducted on January 15, 2025, yielded 105 records; after independent dual-reviewer screening (Cohen's $\kappa = 0.82$ on $n = 105$ title/abstract decisions), 12 empirical studies published between 2013 and 2025 were included and analyzed. Given the small corpus and emergent nature of the field, findings are explicitly characterized as exploratory evidence mapping rather than consolidated empirical conclusions. The corpus spans educational contexts from secondary school to teacher training and engages AI types including intelligent tutoring systems, machine learning, natural language processing, and generative models. Results indicate that AI is used in four pedagogical roles: tutor, evaluator, assistant, and analyst, with documented benefits in conceptual understanding, formative feedback, personalization, and self-regulated learning. Alignment with SDG-4-relevant sustainability competencies (critical thinking, problem solving, collaboration, and digital literacy) was identified through a secondary interpretive analysis by the review authors against Education for Sustainable Development competency frameworks (Rieckmann, 2011; UNESCO, 2017; 2024); this alignment is inferential and design-based rather than directly measured by the primary studies. Ongoing challenges include technical reliability, teacher preparation and acceptance, and ethical concerns regarding transparency, bias, equity, and data protection. The review contributes an initial, disciplinary-specific evidence map at the intersection of AI, physics education, and SDG 4, and identifies six evidence gaps for future empirical research. No protocol was preregistered prior to conducting this review; this absence is acknowledged as a transparency limitation.

KEY WORDS: Artificial intelligence, education for sustainable development, physics education, preferred reporting items for systematic reviews and meta-analyses, sustainability competencies, sustainable development goal 4, systematic literature review

INTRODUCTION

The incorporation of emerging technologies has driven significant transformations in the teaching and learning of physics over the past decades (Gualán et al., 2025). Among these innovations, artificial intelligence (AI), defined as the capacity of computer systems to perform tasks that traditionally required human intelligence, such as reasoning, perception, and complex problem solving (Russell and Norvig, 2021), has gained increasing relevance for its ability to create adaptive and personalized learning environments (Almasri, 2024; Zawacki-Richter et al., 2019). In education, AI enables the analysis of learning patterns, delivery of feedback tailored to student needs, and design of interactive experiences supporting the understanding of complex concepts (Almasri, 2024; Holmes and Tuomi, 2022).

In physics education specifically, AI applications are diverse: Intelligent tutoring systems that adjust content according to student progress (Han et al., 2019; Walker et al., 2014);

algorithms for predicting academic performance and enabling early intervention (Zabriskie, 2019); natural language processing tools for automated assessment of open-ended responses (Campbell et al., 2024); and adaptive simulations to enhance understanding of physical phenomena such as kinematics and dynamics (Fang and Guo, 2016). More recently, generative AI models have begun to be evaluated both as tutoring agents and as benchmarks for physics reasoning (Dunlap et al., 2025; Meyer et al., 2025; Polverini and Gregorcic, 2024).

These developments connect directly to Sustainable Development Goal 4 (SDG 4), which seeks to ensure inclusive, equitable, and quality education and promote lifelong learning opportunities for all (UNESCO, 2015). When implemented with ethical and equity-oriented approaches, AI can broaden access to educational resources, adapt teaching processes to student characteristics, and foster sustainability competencies, including critical thinking, problem solving, collaboration, and digital literacy (Boeren, 2019; Rieckmann, 2011; Vinuesa et

al., 2020; UNESCO, 2024). In physics education specifically, SDG-aligned technologies have been linked to gains in conceptual understanding, engagement, and critical thinking, provided that digital divides are addressed and academic integrity is safeguarded (Okulich-Kazarin et al., 2024; Su et al., 2023).

Although the number of studies in this field has increased in recent years, the available literature reflects considerable diversity in methodological approaches, AI types, competencies assessed, and educational levels addressed. This heterogeneity complicates the identification of common patterns and the assessment of contributions to SDG 4 advancement. Existing systematic reviews on AI in education (Almasri, 2024; Chiu et al., 2023; Su et al., 2026; Zawacki-Richter et al., 2019) have mapped general trends but do not specifically address physics as a discipline, nor do they examine the articulation between AI-mediated pedagogical strategies and Education for Sustainable Development (ESD) competency frameworks within the SDG 4 context.

The originality of the present review lies in the joint articulation of three dimensions not previously examined in combination: (a) disciplinary specificity: Focusing on physics education rather than STEM or science broadly; (b) competency mapping: Systematically applying established ESD frameworks (Rieckmann, 2011; UNESCO, 2017, 2024) to AI-mediated pedagogical designs; and (c) SDG 4 alignment: Evaluating the extent to which reported practices contribute to quality, inclusive, and equitable education as defined by the 2030 Agenda. Given the small corpus ($n = 12$) and the emergent nature of the field, this work is characterized as an exploratory systematic review; all findings should be read as an initial evidence map and hypothesis-generating synthesis rather than consolidated or generalizable conclusions.

Three methodological cautions guide the interpretation of this review. First, the sustainability competency mapping presented in the results section reflects interpretations made by the review authors rather than outcomes directly measured in the primary studies. Second, the use of a single database limits the completeness of the evidence base, particularly for research from emerging academic communities. Third, the absence of protocol preregistration represents a limitation in methodological transparency. Together, these considerations bound the scope of the findings.

Against this background, the present study addresses the following research question: What types of AI have been implemented in physics education, what pedagogical roles do they play, what effects on learning have been reported, and how do these applications align, through an inferential and design-based analysis, with the development of sustainability-related competencies within the framework of SDG 4?

The general objective is to systematically analyze the scientific literature on AI in the teaching and learning of physics, focusing on AI types and roles, effects on learning, and their

plausible contributions to sustainability-related competencies in the context of SDG 4. Four specific objectives (SO) organize the analysis:

- SO1. Describe trends in scientific production on AI in physics education (publication geography, temporality, journal distribution, methodological diversity).
- SO2. Identify AI types, pedagogical roles, implementation frameworks, and theoretical foundations used in educational experiences in physics.
- SO3. Examine how AI-mediated pedagogical strategies may align with key sustainability competencies within the framework of SDG 4, distinguishing between explicit alignment identified in the reported instructional design and plausible alignment inferred by the review authors. Neither category is considered a direct empirical measurement of ESD competency outcomes.
- SO4. Identify benefits, limitations, recommendations, and challenges reported by the reviewed studies regarding AI integration in physics education.

REFERENTIAL FRAMEWORK

AI in Education: Roles, Frameworks, and Challenges

AI has gained an increasingly prominent role in education due to its capacity to create learning environments tailored to individual student characteristics (Chen et al., 2022; Kamalov et al., 2023). Its development has evolved from tutorial systems focused on individualized content delivery toward more sophisticated platforms integrating data analysis, natural language processing, predictive modeling, and interactive experiences (Lameras and Arnab, 2022), accompanied by pedagogical approaches emphasizing personalization and flexibility (Zawacki-Richter et al., 2019). The literature reports applications ranging from adaptive tutoring systems and automated assessment tools to conversational virtual assistants, gamified environments, and learning analytics (Baker et al., 2019; Campbell et al., 2024; Holmes and Tuomi, 2022).

Implementation presents challenges related to algorithmic transparency, data protection, bias mitigation, and equity of access (Nguyen et al., 2023). In response, international organizations have proposed frameworks incorporating technical, reflective, and ethical dimensions to guide safe and inclusive AI integration in education (UNESCO, 2024). Table 1 summarizes these dimensions with concrete educational application examples that extend the original framework.

AI in Physics Education and Its Link to SDG 4

Building on these general frameworks, physics education offers particularly diverse opportunities for AI integration across specific pedagogical goals (Almasri, 2024). Effectiveness depends not only on the technology employed (Zawacki-Richter et al., 2019) but also on its integration into instructional sequences (Holmes and Tuomi, 2022), degree of personalization (Han et al., 2019), and the competencies it targets (Rieckmann, 2011). Among the most documented applications are intelligent tutoring systems, which adapt content, difficulty,

Table 1: Technical, reflective, and ethical dimensions of AI applied to education (adapted from Nguyen et al., 2023; UNESCO, 2024), with educational application examples

Dimension	Description	Educational application examples
Technical	Knowledge and skills required to understand, operate, and develop AI tools, including their programming, functioning, and limitations.	Intelligent tutoring system design; NLP-based grading algorithms; adaptive simulation architecture.
Reflective	Ability to critically analyse the role of AI in teaching and learning, considering pedagogical, social, and cultural impact.	Classroom audits of AI-generated feedback; evaluating assumptions in predictive models; discussion of AI limitations.
Ethical	Principles and practices ensuring responsible AI use: transparency, equity, privacy, and bias minimization.	Algorithmic transparency protocols; equity-auditing of adaptive systems; data-protection compliance in student analytics.

AI: Artificial intelligence

and feedback according to student progress (Han et al., 2019; Walker et al., 2014), and adaptive simulations enabling the representation of complex physical phenomena through real-time data manipulation (Fang and Guo, 2016; Gualán et al., 2025). Empirical evidence supports these effects: Han et al. (2019) identified improvements in kinematics retention, Fang and Guo (2016) demonstrated enhanced understanding of dynamic phenomena through adaptive simulation, and Gualán et al. (2025) reported increased participation through virtual environments.

When embedded in thoughtful pedagogical designs, AI can improve conceptual understanding (Han et al., 2019; Fang and Guo, 2016), enhance motivation through adaptive simulations (Henze et al., 2024), and strengthen participation through collaborative platforms with immediate feedback (Campbell et al., 2024; Gualán et al., 2025). Its educational impact, however, depends critically on instructional design quality (Holmes and Tuomi, 2022), alignment with learning objectives (Zawacki-Richter et al., 2019), and teacher preparation. Beyond performance and motivation, AI can serve as a bridge toward transversal sustainability competencies (Rieckmann, 2011; UNESCO, 2024); connecting scientific training with civic engagement and SDG 4 aims (Vinuesa et al., 2020). However, these contributions should generally be understood as plausible design affordances rather than demonstrated outcomes of the studies included in the review.

SDG 4 (UNESCO, 2015) aims to ensure inclusive, equitable, and quality education for all. AI can contribute through personalization, adaptive tutoring, and expanded resource access (Zawacki-Richter et al., 2019; Vinuesa et al., 2020), while requiring careful attention to algorithmic transparency, data protection, bias mitigation, and equitable access (Holmes and Tuomi, 2022; Nguyen et al., 2023). Vinuesa et al. (2020) further document the dual-use nature of AI technologies: Capable of both advancing and undermining SDG progress depending on implementation contexts, a tension that runs through the SO4 discussion.

The development of AI in physics education can be analytically periodized into four phases. An initial phase (circa 2000–2015) was characterized by intelligent tutoring systems and personalized content delivery (Han et al., 2019; Walker et al., 2014). An intermediate phase (circa 2015–2020) saw learning analytics and predictive modeling aimed at early intervention

(Zabriskie, 2019). A recent phase (2020–2025), most directly reflected in this review’s corpus, has been defined by NLP, deep learning, and generative models for automated assessment, formative feedback, and disciplinary benchmarking (Campbell et al., 2024; Meyer et al., 2025; Polverini and Gregorcic, 2024). Looking ahead, a future phase, identified from contextual literature rather than the corpus, appears oriented toward governance, ethics, equitable access, and authentic assessment of sustainability competencies (Nguyen et al., 2023; UNESCO, 2024). This periodization reveals a trajectory from tool-centric applications toward pedagogically integrated and ethically aware approaches, which frames the rationale for this review.

Key Sustainability Competencies for Physics Education

The key competencies for sustainability defined by Rieckmann (2011) and adopted by UNESCO (2017; 2024) represent capacities needed to address the social, economic, and environmental challenges of the twenty-first century: Systems thinking, anticipatory competence, normative competence, strategic competence, collaboration, critical thinking, self-awareness, and integrated problem solving. All can plausibly be fostered in physics education through AI integration (Table 2). The mapping presented in Table 2 describes design-based affordances, what may be pedagogically possible, rather than empirically documented outcomes. Future research should assess competency development directly using validated instruments.

In physics education specifically, these competencies enable learners to grasp complex phenomena, evaluate long-term consequences, make informed decisions, and act responsibly in response to global challenges (McComas, 2014). Their development involves not only the acquisition of technical knowledge, but also ethical reflection, collaborative engagement, and collective action aligned with SDG 4. AI can support this process when it is embedded in active, collaborative, and contextualized learning experiences that connect conceptual understanding with transversal competencies. For example, problem-based learning on complex issues allows students to address real-world sustainability challenges, such as energy optimization or climate system analysis, through adaptive simulations and predictive models that enable the exploration of scenarios and consequences (Almasri, 2024; Fang and Guo, 2016). Similarly, scenario simulation and prospective analysis can foster anticipatory competence by allowing learners to evaluate

Table 2: Key sustainability competencies, their definitions, and examples of AI-mediated development in physics education (adapted from Rieckmann, 2011; UNESCO, 2017, 2024)

Competency	Definition (Rieckmann, 2011; UNESCO, 2017; 2024)
Systems thinking	Recognising and understanding relationships within complex natural, social, and technical systems.
Anticipatory	Evaluating future trends and anticipating consequences of present actions and decisions.
Normative	Understanding and negotiating values, principles, and sustainable goals.
Strategic	Designing and implementing collective actions to address sustainability challenges.
Collaborative	Working constructively with others, learning from diverse perspectives and disciplines.
Critical thinking	Questioning norms, practices, and opinions while assessing their validity and consequences.
Self-awareness	Recognising personal limitations, values, and responsibilities in relation to global problems.
Integrated problem-solving	Applying diverse competencies and knowledge to formulate solutions to complex problems.

AI: Artificial intelligence

the long-term impacts of scientific and technological decisions through virtual environments (Gualán et al., 2025). AI-supported pedagogical strategies may also promote normative competence and critical thinking through guided debates and reflective activities, particularly when natural language processing systems are used to analyze the ethical and social implications of physics applications (Campbell et al., 2024). In addition, AI-mediated collaborative work can strengthen collaborative competence through experimental or modeling projects that facilitate task management, data sharing, and immediate feedback (Henze et al., 2024). Finally, intelligent tutoring systems can support metacognition and self-regulation by guiding students in monitoring their progress, identifying areas for improvement, and developing greater self-awareness in their learning process (Han et al., 2019).

METHODOLOGY

This study is a systematic literature review conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021). Its purpose was to identify, analyze, and synthesize empirical research on AI in physics education, with a secondary mapping against SDG 4 and ESD competency frameworks. Although the review followed a systematic procedure, it is characterized as exploratory because the search and eligibility process resulted in a limited final corpus ($n = 12$), consistent with the emergent and heterogeneous nature of the field. No protocol was preregistered before conducting the review; this is acknowledged as a transparency limitation throughout.

Search Strategy and Database Selection

A search was conducted in the Web of Science (WoS) database, selected for its broad coverage of high-impact peer-reviewed journals across science, technology, and education, as well as its support for precise field-level Boolean queries. This database was considered appropriate for an initial evidence-mapping synthesis in a field where a substantial proportion of high-visibility contributions are published in WoS-indexed journals (e.g., Physical Review Physics Education Research, IEEE Transactions on Learning Technologies, Neural Computing and Applications). At the same time, the use of

Table 3: Advanced search equation applied in Web of Science

Search equation

("artificial intelligence" OR "machine learning") AND ("physics education" OR "physics teaching" OR "physics learning")

a single database constitutes a boundary condition of the review. Relevant studies may be indexed in other sources such as Scopus, ERIC, or regional repositories, particularly those representing emerging research communities in Asia and Latin America. Accordingly, the present review should be interpreted as a selective synthesis of WoS-indexed literature rather than a comprehensive mapping of the field. Future reviews would benefit from extending database coverage and including non-English sources to enhance representativeness. Table 3 presents the search equation.

Duplicate records were identified using the WoS built-in deduplication function; manual verification confirmed no remaining duplicates in the 105 retrieved records. This process yielded an initial set of 105 documents.

Screening and Selection

Title/abstract screening was conducted independently by two reviewers (one specializing in physics education; one in educational technology and ESD), using a standardized screening form. Each reviewer applied the eligibility criteria in Table 4 independently. Disagreements were resolved through structured discussion; a third author served as arbiter when bilateral consensus could not be reached. This process yielded Cohen's $\kappa = 0.82$ (calculated across all 105 title/abstract decisions before consensus resolution), indicating strong inter-rater agreement (Landis and Koch, 1977). After abstract screening, 53 records were retained for full-text review. Full-text screening led to the exclusion of 41 records (27 for insufficient disciplinary focus on physics, 8 for lack of empirical design, 4 for being purely technical without educational application, 2 for AI tools not explicitly implemented in an educational context). The final corpus comprised 12 studies.

The selection process is summarized in Figure 1 (PRISMA flow diagram).

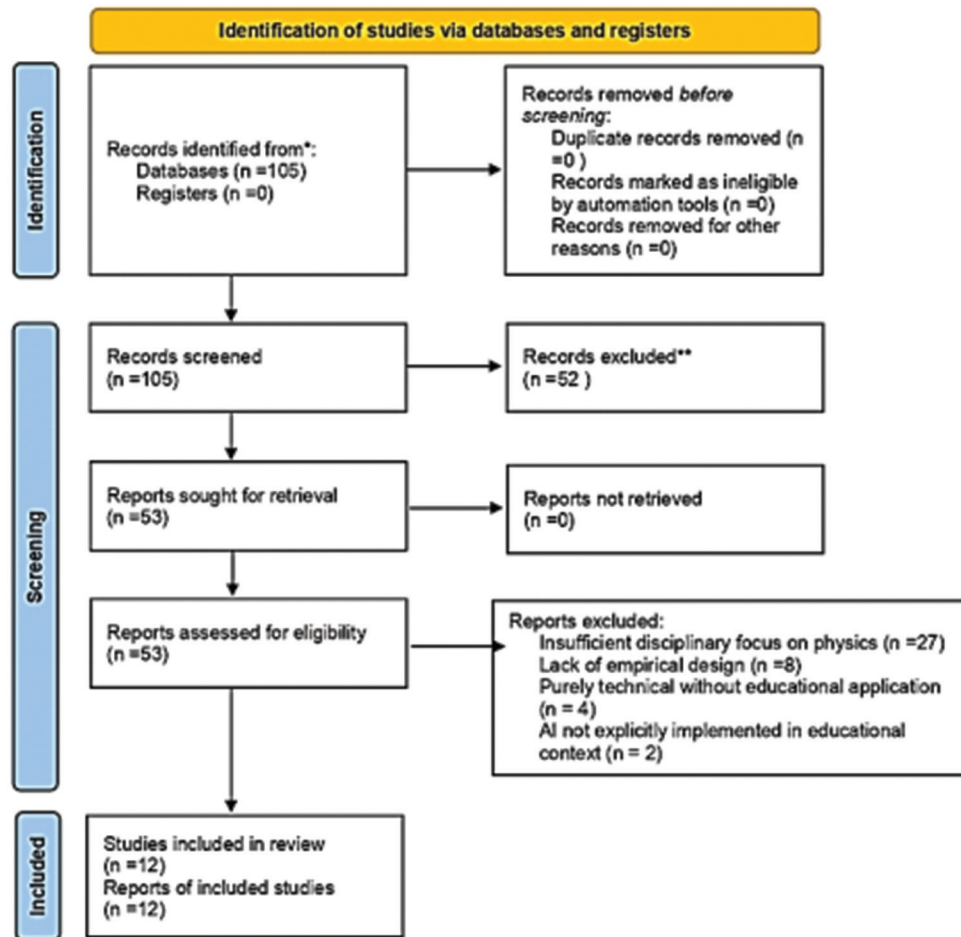


Figure 1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the study selection process

Table 4: Inclusion and exclusion criteria

Variable	Inclusion criteria
Topic	Studies on AI in physics education or learning.
Document type	Journal articles (peer-reviewed).
Language	English or Spanish.
Publication period	No time restriction.
Disciplinary context	Empirical studies on teaching/learning of physics.
Study approach	Empirical research with direct educational application.
AI use	Explicit AI or machine-learning tools in an educational context.
Participants	Students, teachers, teacher educators, or researchers in physics education.

AI: Artificial intelligence

Data Extraction

Once the final corpus was defined, data extraction was organized according to the study's SOs. The included studies were first summarized in terms of basic bibliographic and contextual characteristics, including year, author, title, journal, and country of origin. A common analytical framework was then used to guide extraction across the main categories of analysis. Sustainability competencies

were coded independently by both reviewers using the competency definitions as a codebook. The full text of each study was used as the unit of analysis. To ensure consistency, coding distinguished between explicit design-level alignment identified by the review authors, plausible design-level alignment inferred from the pedagogical design, and insufficient information to infer alignment. Explicit design-level alignment was assigned when a primary study described activities, tasks, or learning contexts that corresponded to a competency definition; this classification did not signify that the competency had been empirically measured. Plausible alignment was assigned when the reported design reasonably suggested a connection but did not explicitly establish or measure the competency. Inter-rater agreement was acceptable, with $\kappa = 0.76$ across 96 coding decisions, and all disagreements were resolved through structured discussion. Table 5 presents the final corpus of included studies, including year of publication, first author, abbreviated title, journal, and country of origin.

After identifying the final corpus, the next step was to define the analytical categories used to examine the studies systematically. Table 6 summarizes these categories, their descriptions, and the corresponding data extraction rules.

Table 5: Final corpus of included studies

ID	Year	1 st author	Title (abbreviated)	Journal	Country
A1	2013	Myneni et al.	ViPS: Intelligent tutoring for physics learning	IEEE Trans. Learning Technologies	USA
A2	2019	Zabriskie	ML to predict physics course outcomes	American Journal of Physics	USA
A3	2023	Demirezen	Deep learning for misconception detection in physics	Neural Computing and Applications	Turkey
A4	2023	Kortemeyer	AI grading of student physics solutions: Feasibility study	Phys. Rev. Physics Educ. Res.	Switzerland
A5	2024	Campbell et al.	IBM Watson NLP as short-answer tool in PER	Phys. Rev. Physics Educ. Res.	USA
A6	2024	Polverini and Gregorcic	ChatGPT performance on kinematics graph test	Phys. Rev. Physics Educ. Res.	Sweden
A7	2025	Wyrwich et al.	Tracking energy understanding progression through AI	Phys. Rev. Physics Educ. Res.	Germany
A8	2025	Megalina et al.	Multimedia AI tool for microteaching physics skills	Jurnal JPPI	Indonesia
A9	2025	Dunlap et al.	LLM descending an inclined plane problem	Phys. Rev. Physics Educ. Res.	USA
A10	2025	Wattanakasiwich et al.	Physics instructors' acceptance of generative AI	Phys. Rev. Physics Educ. Res.	Thailand
A11	2025	Jufrida et al.	AI-driven ethnoscience learning in physics	J. IT Educ.: Innovations in Practice	Indonesia
A12	2025	Meyer et al.	Automatic feedback on physics tasks through open-source GenAI	J. Science Educ. and Technology	Germany

AI: Artificial intelligence

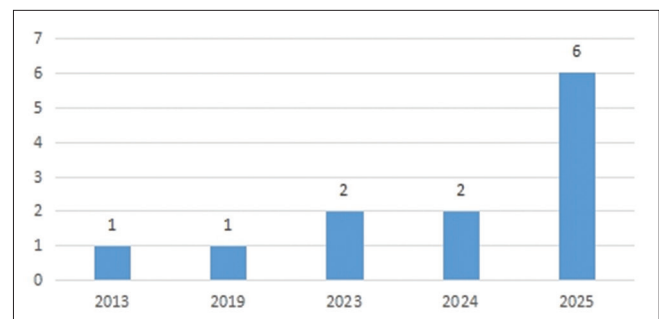
Table 6: Analytical categories, descriptions by specific objective

SO	Category	Description
SO1	Country of first author	Country of first author's institutional affiliation.
	Year of publication	Year the study was published.
	Journal	Name of publication venue.
	Methodological approach	General study design (qualitative, quantitative, mixed, computational).
	Data collection instruments	Techniques/instruments used for empirical data collection.
SO2	AI type and role	AI technology used and pedagogical function (tutor, evaluator, assistant, analyst).
	Theoretical framework	Conceptual/theoretical foundation guiding the study.
SO3	Didactic or methodological model	Presence and type of explicit instructional model.
	Educational level	Educational stage (secondary, university, teacher training).
	Replicability	Sufficiency of procedural description for replication.
	Sustainability competencies	Alignment with Rieckmann (2011)/UNESCO (2017; 2024) competencies.
SO4	Benefits	Reported pedagogical improvements or positive outcomes.
	Recommendations and challenges	Author-stated suggestions, limitations, and warnings.

RESULTS

The results are organized according to the four SOs of the review. With respect to SO1, the results indicate that research on AI in physics education remains limited but has increased notably in recent years. As shown in Figure 2, the final corpus included 12 studies published between 2013 and 2025. Early contributions were scarce, with only one study published in 2013 and another in 2019. From 2023 onward, however, publication activity increased, with two studies published in 2023, two in 2024, and six in 2025. This temporal pattern suggests that AI in physics education is an emerging research area that has gained rapid visibility in the most recent period, particularly alongside the expansion of machine learning, natural language processing, and generative AI applications in educational contexts.

The distribution of publication venues reflects both disciplinary concentration and diversification within the field. A considerable proportion of studies were published in Physical Review Physics Education Research (A2, A4, A5, A6, A7, A9, and A10), while additional contributions appeared in

**Figure 2:** Publication timeline of selected studies (2013–2025)

Journal of Science Education and Technology (A12), Journal of Information Technology Education: Innovations in Practice (A11), Jurnal Penelitian dan Pembelajaran IPA (A8), Neural Computing and Applications (A3), and IEEE Transactions on Learning Technologies (A1). This concentration in physics education journals matters, as these venues apply discipline-specific peer review standards that help ensure alignment between AI-based innovation and physics-pedagogical validity. At the same time, the presence of studies in interdisciplinary

and technology-oriented journals reflects the hybrid and still-expanding nature of the field. Methodological approaches also varied across the corpus. The included studies comprised quantitative experimental designs (A1, A2), case-based and exploratory approaches (A3, A4, and A9), mixed-methods studies involving surveys, structural equation modeling, or content analysis (A5, A10, and A11), a quasi-experimental pre-test–post-test design (A8), and large-scale language model performance comparisons (A6, A7, and A12). Data collection instruments included tests and performance logs (A1, A2), NLP-based clustering and open-response analysis (A3, A5, and A12), automated grading of synthetic or student-generated solutions (A4, A9), digital longitudinal notebooks (A7), teacher surveys (A10), interviews and classroom observations (A11), and expert-validated multimedia materials with questionnaires (A8). This methodological pluralism is a strength in terms of breadth, but it also limits comparability across studies.

With respect to SO2, the results show that the reviewed studies used AI in diverse ways, reflecting different degrees of pedagogical integration. Four main roles were identified across the corpus: Tutor, evaluator, analyst, and assistant. AI was used as a tutor in A1, through a virtual platform with embedded feedback in structured learning sequences, and in A8, through AI-avatar-supported microteaching. It functioned as an evaluator in A3, where deep learning was used for misconception detection; A4, which combined OCR and generative AI for handwritten solution grading; A5, which used NLP for open-response categorization; A6, which examined ChatGPT performance on kinematics graphs; A9, which evaluated GPT-4 on variations of an inclined-plane problem; and A12, which used an open-source LLM for formative feedback. AI was used as an analyst in A2, through machine learning for performance prediction; A7, through machine learning and NLP for tracking learning progressions; and A11, through predictive algorithms embedded in ethnoscience-based projects. Finally, AI was used as an assistant in A10, where generative AI supported teachers in problem generation and instructional material creation.

This distribution suggests that AI in physics education is being used not as a single pedagogical tool, but as a set of heterogeneous functions that operate at different points in the teaching and learning process. Some applications are directly embedded in instructional activity, as in tutoring, feedback, and microteaching contexts. Others are more external or analytical, focusing on prediction, classification, or model performance evaluation. Therefore, the pedagogical significance of AI depends not only on the technology used, but also on the role assigned to it within the instructional design. These roles are not equally pedagogically transformative, a point further developed in the Discussion.

Theoretical frameworks also showed marked heterogeneity. A1 combined sociocultural theory with the Technology Acceptance Model and the Community of Inquiry framework, while A10 drew on Diffusion of Innovations theory and UTAUT2 to

analyze teacher adoption of generative AI. A11 employed culturally responsive pedagogy to support ethnoscience-based physics projects, and A12 emphasized formative assessment and cognitive models of problem solving. A7 combined learning progressions with Evidence-Centered Design for Learning Progression Analytics. In contrast, several studies, particularly A2, A3, and A5, were primarily methodological, privileging technical approaches over explicit pedagogical frameworks. Studies A4, A6, A8, and A9 had less explicit theoretical foundations and framed their contributions mainly in terms of technological feasibility or model performance.

With respect to SO3, the results show considerable variation in the extent to which AI was embedded within explicit pedagogical designs. Some studies incorporated clearly defined instructional approaches. A1 organized learning through a collaborative problem-solving intelligent tutoring system, while A8 and A11 used project-based learning in microteaching and ethnoscience-based physics activities, respectively. A7 also adopted a project-based orientation, integrating it with Evidence-Centered Design for Learning Progression Analytics. In contrast, A2, A3, A4, A5, A6, A9, A10, and A12 did not report a specific didactic model. These studies were mainly oriented toward algorithm evaluation, automated feedback feasibility, model performance, or teacher perception analysis.

The educational contexts were also diverse. Secondary education was represented in A1, A6, A7, A11, and A12; university introductory or teacher education contexts appeared in A2, A3, A4, A5, and A9; and pre-service or in-service teacher training was addressed in A8 and A10. The degree of replicability also varied across the corpus. A1, A5, A7, A8, A11, and A12 provided detailed descriptions of procedures, instruments, or implementation sequences, which makes adaptation to similar contexts more feasible. By contrast, A2, A3, and A9 offered more limited pedagogical detail, as their reporting concentrated mainly on technical procedures or model performance. Participant roles likewise differed across studies: Students acted as active learners in A1, A7, A8, and A11; as respondents whose answers were analyzed by algorithms in A2, A3, A5, A6, A9, and A12; as teacher-participants in A10; and as absent direct participants in A4, where the focus was on AI-evaluated subject matter. For the sustainability competency mapping, three review-based categories were used. Explicit design-level alignment (■) indicates that the review authors identified pedagogical activities, tasks, or learning contexts in the primary study that corresponded to the competency definition; it does not indicate empirical measurement of that competency. Plausible design-level alignment (▣) indicates that the review authors inferred a reasonable connection from the reported instructional design, learning activity, or AI-supported context. Insufficient information (□) indicates that no alignment could be inferred from the available reporting. Under this interpretive coding, A7 and A11 displayed the broadest patterns of design-level alignment. The review authors interpreted A7's 10-week project-based unit on school energy supply, combined with

learning progression analytics, as aligning with systems thinking, anticipatory reasoning, normative judgment, strategic action, critical thinking, collaboration, and integrated problem solving. Similarly, A11's ethnoscience-based physics projects and data-driven personalization were interpreted as aligning with systems thinking, anticipatory reasoning, strategic action, collaboration, and integrated problem solving, while offering plausible design pathways for normative competence, critical thinking, and self-awareness.

Other studies offered more conditional or context-dependent design affordances. The review authors inferred that A8's AI-avatar-supported microteaching created possible pathways for normative competence, self-awareness, collaboration, strategic action, and integrated problem solving, particularly when projects are framed around sustainability-related challenges. A10's focus on ethics, privacy, linguistic equity, and responsible AI use was interpreted as opening space for critical, normative, anticipatory, and systems-oriented reflection in teacher education. A12 provided a design context from which strategic action, integrated problem solving, and self-awareness could plausibly be inferred through the auditing of AI-generated feedback. In contrast, the reported designs of A2 and A3 did not provide sufficient information to infer sustainability competency alignment without additional instructional orchestration. Table 7 presents the complete review-based competency matrix.

Quality and Methodological Characteristics of Included Studies

A structured narrative quality assessment was conducted across four dimensions for all included studies. Each dimension was rated as High (H), Moderate (M), or Low (L). A High rating indicates that the study provided clear, detailed, and sufficient information for that dimension; a Moderate rating indicates that the information was generally adequate but lacked some level of detail or specificity; and a Low rating indicates that the information was limited, incomplete, or insufficient to support

a robust assessment. Design clarity was rated High across all 12 studies, reflecting consistently clear descriptions of objectives and procedures. Sample description was high in A7, A10, and A11, due to detailed participant and selection reporting, and Low in A3, A6, and A9, due to limited demographic or sample size information, with the remaining studies rated Moderate. Outcome measurement was High in A1, A5, A7, A11, and A12, and Moderate or Low in others, particularly those evaluating language model performance on limited test sets or relying mainly on perception data. Replicability was high in studies providing detailed implementation sequences and Moderate in technically focused studies. These heterogeneities reinforce the importance of cautious interpretation and the need for more rigorous designs in future research. Table 8 presents the full quality matrix.

With respect to SO4, the reviewed studies report several benefits associated with the use of AI in physics education, particularly in relation to conceptual understanding, problem solving, feedback, and instructional efficiency. In A1, the intelligent tutoring system supported collaborative knowledge construction through structured guidance and feedback. A7 showed how machine learning and natural language processing can support the mapping of students' learning progressions in energy-related concepts, while A12 highlighted the contribution of open-source language models to the quality and timeliness of formative feedback on physics tasks. Together, these studies suggest that AI can contribute to physics learning when it is meaningfully connected to instructional sequences and formative assessment processes.

Other benefits were related to efficiency, personalization, and the diversification of teaching resources. A5 reported that NLP-based categorization of open responses reduced the time required for analysis; while A10 showed that generative AI could support teachers in creating physics problems and instructional materials, potentially reducing workload. In teacher education, A8 found that AI avatars increased motivation and improved

Table 7: Sustainability competency alignment inferred by the review authors (SO3): ■ = explicit design-level alignment identified from the primary study; ▣ = plausible design-level alignment inferred from the pedagogical design; □ = insufficient information to infer alignment. Symbols do not represent direct empirical measurement of ESD competencies

ID	Reference	Sys. Think.	Anticip.	Normat.	Crit. Think.	Strategic	Collab.	Integr. Prob.	Self- Aware.
A1	Myneni et al. (2013)	□	□	□	■	■	■	□	■
A2	Zabriskie (2019)	□	□	□	□	□	□	□	□
A3	Demirezen (2023)	□	□	□	▣	□	□	▣	□
A4	Kortemeyer (2023)	□	□	□	▣	□	□	▣	□
A5	Campbell et al. (2024)	□	□	□	▣	□	□	▣	■
A6	Polverini and Gregorcic (2024)	□	▣	▣	■	▣	□	■	□
A7	Wyrwich et al. (2025)	■	■	■	■	■	■	■	□
A8	Megalina et al. (2025)	□	□	▣	□	■	■	■	▣
A9	Dunlap et al. (2025)	□	□	□	□	■	□	▣	□
A10	Wattanawasiwich et al. (2025)	▣	▣	▣	■	■	□	▣	□
A11	Jufrida et al. (2025)	■	■	▣	▣	■	■	■	▣
A12	Meyer et al. (2025)	□	□	□	□	■	□	■	▣

ESD: Education for Sustainable Development

Table 8: Summary quality matrix of included studies (H=High; M=Moderate; L=Low)

ID	References	Design clarity	Sample description	Outcome measurement	Replicability
A1	Myneni et al. (2013)	H	M	H	H
A2	Zabriskie (2019)	H	M	M	M
A3	Demirezen (2023)	H	L	M	M
A4	Kortemeyer (2023)	H	M	M	H
A5	Campbell et al. (2024)	H	M	H	H
A6	Polverini and Gregorcic (2024)	H	L	M	M
A7	Wyrwich et al. (2025)	H	H	H	H
A8	Megalina et al. (2025)	H	M	M	M
A9	Dunlap et al. (2025)	H	L	M	M
A10	Wattanakasiwich et al. (2025)	H	H	M	H
A11	Jufrida et al. (2025)	H	H	H	H
A12	Meyer et al. (2025)	H	M	H	H

access to microteaching support materials. Similarly, A11 showed that personalized AI-supported activities strengthened student engagement and helped connect physics learning with local cultural practices. These findings indicate that AI may support not only student learning, but also instructional design, teacher work, and contextualized pedagogy. A further group of studies examined AI model performance in comparison with students. A6 reported that multimodal language models achieved performance comparable to secondary students on kinematics graph tasks, while A9 found that GPT-4 outperformed students on variations of a classical mechanics problem. These results suggest that language models may have potential as practice, feedback, or assessment-support tools. However, their educational value depends on how they are integrated into learning activities and on whether students are guided to critically evaluate AI-generated responses rather than simply accept them.

The recommendations reported across the corpus converge on the need for teacher supervision, critical AI use, and responsible implementation. A2 emphasized that predictive systems should support, not replace, teacher judgment in identifying students who may require additional support. A3 and A4 warned that current algorithmic limitations make premature reliance on automated grading or diagnostic systems problematic. Similarly, A9 stressed the importance of helping students recognize assumptions, inconsistencies, and possible contradictions in LLM-generated solutions. These recommendations suggest that AI should be treated as a pedagogical support system rather than an autonomous substitute for professional judgment.

Teacher preparation and interdisciplinary collaboration also emerged as central conditions for meaningful implementation. A10 recommended specific training in prompt engineering, digital literacy, and responsible generative AI use, while A11 and A12 emphasized collaboration between physics education specialists and AI experts to ensure that digital resources are both pedagogically sound and technically reliable. Equity-related concerns were also recurrent. A10 raised issues of linguistic equity and computational energy costs, while A11

highlighted the importance of culturally responsive AI design. In line with broader concerns raised by Nguyen et al. (2023) and Vinuesa et al. (2020), these findings indicate that AI integration aligned with SDG 4 requires not only technical effectiveness, but also attention to access, cultural relevance, transparency, and fairness.

DISCUSSION

The discussion is organized according to the four SOs of the review. With respect to SO1, the publication pattern identified in this review suggests that AI in physics education is emerging as a recognizable but still weakly consolidated line of inquiry. Although the increase in studies from 2023 onward indicates growing interest, this growth should be interpreted cautiously. Because the search was restricted to WoS, the findings represent an exploratory evidence map of high-visibility WoS-indexed literature rather than a comprehensive synthesis of the global literature on AI in physics education. Compared with broader reviews of AI in education and science education, the physics-specific corpus remains small and heterogeneous (Almasri, 2024; Chiu et al., 2023; Zawacki-Richter et al., 2019). This suggests that physics education has not yet developed a cumulative evidence base comparable to the wider AI-in-education field. This limited consolidation is relevant because the expansion of AI research does not necessarily imply pedagogical maturity. The evidence reviewed here remains distributed across diverse countries, journals, and methodological approaches, with uneven theoretical and instructional grounding. This pattern is consistent with broader concerns in AI education research, where technological innovation often advances faster than evidence on instructional design, teacher mediation, and learning value (Holmes and Tuomi, 2022; Zawacki-Richter et al., 2019). Therefore, the contribution of SO1 is not merely to show an increase in publications, but to identify a field in transition. AI in physics education is becoming more visible, yet it remains geographically and methodologically limited. This is especially important for SDG 4, because claims about inclusive and equitable AI-supported education require

evidence from diverse educational systems and contexts. As Boeren (2019) emphasizes, educational development should not be assessed only by expansion, but by its capacity to generate meaningful and equitable impact. Thus, consolidating a disciplinary research identity, grounded in physics education problems, stronger methodological comparability, and broader geographical representation, matters more than simply producing more studies.

With respect to SO2, the corpus reveals substantial technological diversity, but this diversity should not be equated with pedagogical depth. AI was used as tutor, evaluator, assistant, and analyst; however, these roles differ in their educational significance. Across the reviewed studies, AI was more frequently used to automate feedback, classify responses, predict performance, or benchmark model outputs than to transform instructional design in sustained ways. This pattern is consistent with broader AIED literature, which shows that AI applications often advance technically before their pedagogical integration is fully established (Holmes and Tuomi, 2022; Luckin et al., 2016; Su et al., 2026). The presence of multiple AI types therefore indicates an expanding field, but not necessarily a mature one. What remains less clear is under which instructional conditions these tools become meaningful for disciplinary learning in physics. The heterogeneity observed in this review also reflects the multifunctionality of AI described by Holmes and Tuomi (2022) and Hwang et al. (2020). However, the lack of comparable pedagogical frameworks limits systematic assessment across studies. Future research should therefore move beyond performance outputs and examine how AI affects learning processes, teacher mediation, and the organization of physics instruction (Luckin et al., 2016).

With respect to SO3, the findings suggest that the strongest potential for sustainability-oriented learning does not arise from AI use itself, but from the way AI is embedded within context-rich instructional designs. Studies that integrate AI into project-based, progression-oriented, or decision-focused sequences appear more educationally meaningful than those where AI is used as a peripheral tool or *post hoc* analytic resource. This supports the idea that pedagogy-first integration is essential if AI is to contribute to SDG 4 and to the development of sustainability-related capacities in physics education (Castañeda and Selwyn, 2018; Chen et al., 2022; Holmes and Tuomi, 2022). The distribution of studies across secondary, university, and teacher education contexts also suggests that AI-supported sustainability learning should not be limited to higher education. If competencies such as systems thinking, anticipatory reasoning, collaboration, and integrated problem solving are to be developed progressively, they need to be addressed across educational stages, in line with UNESCO's ESD framework (UNESCO, 2017). Most included studies, however, did not directly assess sustainability competencies as defined by ESD frameworks. Consequently, Table 7 presents a secondary interpretive analysis rather than direct empirical evidence of competency development. All

classifications, including the ■ category, describe alignment identified in the reported pedagogical design and should not be interpreted as primary-study measurements. This distinction prevents overstatement and clarifies that the review identifies plausible educational affordances rather than demonstrated competency outcomes. Future studies should develop and validate instruments capable of directly assessing sustainability competencies in AI-supported physics learning environments.

With respect to SO4, the evidence base is better understood as conditional rather than conclusive. The reported benefits, including more immediate feedback, improved error detection, support for self-regulated learning, and personalization, suggest that AI can add educational value in physics contexts. However, these benefits remain largely context-bound, tool-specific, and dependent on teacher mediation. In this sense, the current corpus points to promising uses of AI, but not yet to a stable or generalizable evidence base for effectiveness (Almasri, 2024; Zawacki-Richter et al., 2019). The limitations identified across the studies should therefore be treated as central challenges rather than secondary implementation issues. Concerns related to validity, model accuracy, teacher acceptance, transparency, bias, privacy, and unequal access define the conditions under which AI can contribute responsibly to physics education. This is particularly relevant for SDG 4, because the same technologies that may support inclusion and personalization can also reproduce inequalities if introduced without governance, teacher preparation, or equity safeguards (Holmes and Tuomi, 2022; Vinuesa et al., 2020). Overall, AI remains a dual-use technology whose educational contribution depends on how, where, and for whom it is implemented (Vinuesa et al., 2020).

The synthesis of SO1–SO4 identifies six evidence gaps that limit the field's capacity to move from exploratory mapping toward stronger empirical claims. Table 9 connects each gap to an actionable research priority: (1) develop and validate instruments for directly assessing all eight ESD competencies in AI-supported physics environments, using pre-test-post-test measures and control groups; (2) conduct randomized or quasi-experimental comparisons across secondary, university, and teacher-education levels; (3) broaden cultural and geographic representation by extending database coverage and supporting studies in underrepresented regions; (4) implement controlled longitudinal interventions lasting at least one academic year to examine sustained conceptual, motivational, and competency outcomes; (5) integrate equity, access, linguistic representation, and algorithmic-bias audits into empirical designs; and (6) evaluate teacher professional development, instructional-design capacity, and adoption barriers through mixed methods. In particular, controlled longitudinal interventions that directly assess ESD competencies are needed to replace short-term, design-based inference with stronger evidence of educational impact. These recommendations identify priorities for future primary research rather than issues resolved by the present review.

Table 9: Six evidence gaps identified from the review and recommendations for future empirical research

Gap ID	Evidence gap	Description	Future research recommendation
Gap 1	ESD competency assessment instruments	No included study directly measured sustainability competencies using validated instruments.	Develop and validate assessment tools for all 8 ESD competencies in physics AI environments; include pre/post designs with control groups.
Gap 2	Controlled comparative studies across educational levels	The corpus spans secondary, university, and teacher-training contexts but lacks comparative designs across levels.	Conduct cross-level quasi-experimental or randomized studies examining how pedagogical impact of AI varies by educational stage.
Gap 3	Cultural and geographic diversity in corpus	Evidence is concentrated in North America, Europe, and Southeast Asia; Latin American and African contexts are absent.	Expand search to Scopus, ERIC, and regional repositories; commission or support research in underrepresented regions.
Gap 4	Longitudinal learning impact beyond single sessions	Studies report short-term or task-specific outcomes; sustained effects on conceptual development remain undocumented.	Design controlled longitudinal intervention studies (minimum one academic year), using comparison groups where feasible, to track conceptual understanding, motivation, and directly assessed competency development.
Gap 5	Equity, access, and bias auditing	Equity conditions (digital access, algorithmic bias, linguistic representation) are raised but not empirically investigated in corpus studies.	Integrate equity audits into AI-physics studies; examine differential outcomes across socioeconomic, gender, and linguistic groups.
Gap 6	Teacher professional development and adoption	A single study (A10) examined teacher acceptance; instructional design competencies for AI integration remain uninvestigated.	Conduct mixed-methods studies on teacher preparation, professional learning, and adoption barriers for AI in physics education.

The gaps are ordered according to their relevance in the SO1–SO4 synthesis. They should be understood as priorities for future empirical research, not as issues resolved by the present review. ESD: Education for Sustainable Development

Limitations

Several limitations constrain the scope and generalizability of this review. First, the search was restricted to WoS, providing selective coverage of high-visibility evidence but excluding work indexed exclusively in Scopus, ERIC, Dimensions, or regional repositories, particularly from emerging research communities in Asia and Latin America. Second, the search equation may not have retrieved studies using alternative terminologies (e.g., deep learning, neural networks, or educational data mining without AI), representing a further source of incompleteness. Third, the final corpus is small ($n = 12$), limiting generalizability and precluding meta-analytic conclusions; all findings should be interpreted as an initial qualitative synthesis. Fourth, no formal quality appraisal or risk-of-bias assessment was applied; as explained in the methodology, this was a deliberate decision justified by design heterogeneity and descriptive scope, with a narrative quality assessment and summary matrix (Table 8) provided as a partial alternative. Fifth, the coding of sustainability competencies in Table 7 is interpretive and inferential; the included studies did not in general directly assess competencies as defined in ESD frameworks, and ratings reflect review authors' analytical judgment rather than observed outcomes. Sixth, the review is susceptible to publication bias. Seventh, given the rapid evolution of AI technologies, findings concerning generative models may quickly become outdated. Eighth, no protocol was preregistered, constituting a transparency limitation. Future reviews should expand database coverage, include non-English sources, apply formal quality appraisal tools (e.g., AMSTAR 2 for systematic reviews; CASP for primary studies), pre-register the protocol, and extend to longitudinal designs.

CONCLUSION

AI is becoming an increasingly visible presence in physics education, spanning applications from intelligent tutoring systems to generative models used for feedback, assessment, and problem-solving support. Across the reviewed studies, AI shows potential to enhance conceptual understanding, personalization, and forms of learner support. At the same time, the educational significance of these applications appears to depend less on the technology itself than on the pedagogical conditions under which it is implemented. A central contribution of this review lies in clarifying that the most plausible links between AI in physics education and SDG 4 arise when AI is embedded within context-rich, decision-oriented instructional designs rather than used as a peripheral analytic or technical add-on. The review contributes not only by mapping existing applications, but also by specifying the instructional and ethical conditions under which AI may more meaningfully support quality, inclusive, and sustainability-oriented physics education, and by identifying six priority evidence gaps (Table 9) for future empirical research.

These conclusions should be interpreted cautiously because the available evidence remains limited and heterogeneous. The reported alignment with sustainability competencies is largely inferential and derives from the review authors' analysis of pedagogical designs rather than from competencies directly assessed in the primary studies. The review should therefore be understood as an initial evidence map that summarizes current research directions while clearly acknowledging the limits of the available evidence. Progress in this field will require moving beyond tool-centered adoption toward pedagogy-first implementation; developing validated assessment instruments for ESD competencies in physics AI learning environments;

conducting controlled comparative studies across educational levels and cultural contexts; strengthening methodological transparency through protocol preregistration; and extending database coverage to Scopus, ERIC, and regional repositories. Equity dimensions, including access, linguistic representation, and socioeconomic context, must be prioritized to ensure that AI adoption supports rather than undermines the inclusive aspirations of SDG 4.

ETHICS APPROVAL

Not applicable. The study does not involve any human participants.

INFORMED CONSENT

Not applicable.

FUNDING

Not applicable. This research received no specific funding from public, commercial, or not-for-profit funding agencies.

AUTHORS' CONTRIBUTIONS

All authors actively participated in conceptualization, data collection and analysis, drafting and critical revision of the manuscript, and approved the final version.

COMPETING INTERESTS

The authors declare no competing interests.

AVAILABILITY OF DATA AND MATERIALS

The full PRISMA checklist and screening forms are available from the corresponding author upon request. The search equation and inclusion/exclusion criteria are reported in the manuscript.

PROTOCOL REGISTRATION

No protocol was preregistered prior to conducting this review. This is acknowledged as a transparency limitation.

REFERENCES

- Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education*, 54(3), 977-997.
- Baker, T., Smith, L., & Anissa, N. (2019). *Educ-AI-tion Rebooted? Exploring the Future of Artificial Intelligence in Schools and Colleges*. NESTA. Available from: <https://www.nesta.org.uk/report/education-rebooted> [Last accessed on 2026 Mar 20].
- Boeren, E. (2019). Understanding sustainable development goal (SDG) 4 on "quality education" from micro, meso and macro perspectives. *International Review of Education*, 65(2), 277-294.
- Campbell, J., Ansell, K., & Stelzer, T. (2024). Evaluating IBM's Watson natural language processing artificial intelligence as a short-answer categorization tool for physics education research. *Physical Review Physics Education Research*, 20(1), 010116.
- Castañeda, L., & Selwyn, N. (2018). More than tools? Making sense of the ongoing digitizations of higher education. *International Journal of Educational Technology in Higher Education*, 15, 22.
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. *Educational Technology and Society*, 25(1), 28-47.
- Chiu, T.K., Xia, Q., Zhou, X., Chai, C.S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education Artificial Intelligence*, 4, 100118.
- Demirezen, M. (2023). Using deep learning to identify student misconceptions in physics problem solving. *Neural Computing and Applications*, 35, 19823-19836.
- Dunlap, J.C., Sissons, R., & Widenhorn, R. (2025). Descending an inclined plane with a large language model. *Physical Review Physics Education Research*, 21(1), 010153.
- Fang, N., & Guo, Y. (2016). Interactive computer simulation and animation for improving student learning of particle kinetics. *Journal of Computer Assisted Learning*, 32(5), 443-455.
- Gualán, J.C.A., Veas, L.J.V., Poveda, A.G.T., Zambrano, N.M.M., Valarezo, S.N.E., & Jiménez, E.L.J. (2025). Tecnologías emergentes en la enseñanza de la física: Realidad virtual y simulaciones educativas. *South Florida Journal of Development*, 6(3), e5023.
- Han, J., Zhao, W., Jiang, Q., Oubibi, M., & Hu, X. (2019). Intelligent tutoring system trends 2006-2018: A literature review. In: *2019 Eighth International Conference on Educational Innovation through Technology (EITT)*. United States: IEEE, p153-159.
- Henze, J., Bresges, A., & Becker-Genschow, S. (2024). *AI-Supported Data Analysis Boosts Student Motivation and Reduces Stress in Physics Education*. [arXiv Preprint].
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570.
- Hwang, G.J., Xie, H., Wah, B.W., & Gasevic, D. (2020). Vision, challenges, roles, and research issues of artificial intelligence in education. *Computers and Education Artificial Intelligence*, 1, 100001.
- Jufrida, J., Kurniawan, W., Furqon, F., Anwar, K., Falah, H.S., & Riantoni, C. (2025). AI-driven ethnoscience learning: Enhancing physics education through Malay cultural insights. *Journal of Information Technology Education Innovations in Practice*, 24, 13.
- Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, 15(16), 12451.
- Kortemeyer, G. (2023). Toward AI grading of student problem solutions in introductory physics: A feasibility study. *Physical Review Physics Education Research*, 19(2), 020163.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174. <https://doi.org/10.2307/2529310>
- Lameras, P., & Arnab, S. (2022). Power to the teachers: An exploratory review on artificial intelligence in education. *Information*, 13(1), 14.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L.B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. London: Pearson Education.
- McComas, W.F. (2014). *The Language of Science Education: An Expanded Glossary of Key Terms and Concepts in Science Teaching and Learning*. Netherlands: Sense Publishers.
- Megalina, Y., Rugaya, R., Amdani, K., & Sihombing, R.A. (2025). Development of a multimedia learning tool by utilizing artificial intelligence (AI) towards basic teaching skills in microteaching courses for physics learning. *Jurnal Penelitian dan Pembelajaran IPA*, 11(1), 113-131.
- Meyer, A., Bleckmann, T., & Friege, G. (2025). Automatic feedback on physics tasks using open-source generative artificial intelligence. *International Journal of Science Education*, 1-26.
- Myneni, A., Rebello, S., & Bouck, E. (2013). Design and evaluation of ViPS, a virtual physics simulation environment: Intelligent tutoring for physics learning. *IEEE Transactions on Learning Technologies*, 6(3), 216-227.
- Nguyen, A., Ngo, H.N., Hong, Y., Dang, B., & Nguyen, B.P.T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), 4221-4241.
- Okulich-Kazarin, V., Artyukhov, A., Skowron, Ł., Artyukhova, N.,

- Dluhopolskyi, O., & Cwynar, W. (2024). Sustainability of higher education: Study of student opinions about the possibility of replacing teachers with AI technologies. *Sustainability*, 16(1), 55.
- 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., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
- Polverini, G., & Gregoric, B. (2024). Performance of ChatGPT on the test of understanding graphs in kinematics. *Physical Review Physics Education Research*, 20(1), 010109.
- Rieckmann, M. (2011). Future-oriented higher education: Which key competencies should be fostered through university teaching and learning? *Futures*, 44(2), 127-135.
- Russell, S.J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. 4th ed. London: Pearson.
- Su, C.S., Díaz-Levicoy, D., Vásquez, C., & Hsu, C.C. (2023). Sustainable development education for training and service teachers teaching mathematics: A systematic review. *Sustainability*, 15(10), 8435.
- Su, C.S., Hsu, C.C., & González Campos, J.A. (2026). ChatGPT as a resource to strengthen mathematics teaching and learning: A systematic review. *Journal of Computer Assisted Learning*, 42(3), e70241.
- UNESCO. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. UNESCO. Available from: <https://sustainabledevelopment.un.org/post2015/transformingourworld> [Last accessed on 2026 Mar 20].
- UNESCO. (2017). *Education for the Sustainable Development Goals: Learning objectives*. UNESCO. Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000247444> [Last accessed on 2026 Mar 20].
- UNESCO. (2019). *Global Education Monitoring Report 2019: Migration, Displacement and Education-Building Bridges, not Walls*. UNESCO. Available from: <https://www.unesco.org/gem-report/en/publication/migration-displacement-education-building-bridges-not-walls> [Last accessed on 2026 Mar 20].
- UNESCO. (2024). *AI Competency Framework for Students*. UNESCO. Available from: <https://www.unesco.org/en/articles/ai-competency-framework-students> [Last accessed on 2026 Mar 20].
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M., & Nerini, F.F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11, 233.
- Walker, E., Rummel, N., & Koedinger, K.R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, 24(1), 33-61.
- Wattanakaswich, P., Kaewkhong, K., & Katwibun, D. (2025). Physics instructors' acceptance and implementation of generative AI. *Physical Review Physics Education Research*, 21(1), 010155.
- Wyrwich, T., Kubsch, M., Drachsler, H., & Neumann, K. (2025). Tracking students' progression in developing understanding of energy using AI technologies. *Physical Review Physics Education Research*, 21(1), 010152.
- Zabriskie, C. (2019). Using machine learning and educational data mining to predict student performance in physics. *Physical Review Physics Education Research*, 15(2), 020110.
- Zawacki-Richter, O., Marín, V.I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education-Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39.