Review Article

AI-Assisted science lesson planning for innovative pedagogy in resource-constrained settings: A scoping review

Received: 30/06/26
Revised: 28/09/26
Accepted: 29/09/26
Published:30/09/26

Ezechiel Nsabayezu[1]*, Gedeon Habakubaho[2], Fidele Mweruli[3]

[1]Department of Mathematics and Science Education, INES-Ruhengeri, Musanze, Rwanda, P.O.B 155 Musanze
[2]Department of Land Survey, INES-Ruhengeri, Musanze, Rwanda, P.O.B 155 Musanze
[3]Department of Sciences and Postgraduate Diploma in Education (PGDE), Kibogora Polytechnic (KP), Nyamasheke, Rwanda

Address for Correspondence
Nsabeyezu (e.nsabayezu@ines.ac.rw)

Abstract

Science teachers, particularly those working in resource-constrained settings, face persistent challenges related to limited laboratory facilities, large class sizes, inadequate instructional materials, and the substantial time required for lesson preparation and resource development. The emergence of generative artificial intelligence (AI), including large language models such as ChatGPT, Gemini, Claude, and Copilot has created new opportunities to support science teachers in preparing lessons and developing instructional resources. However, evidence on how AI is being used for science lesson planning, its reported educational benefits, and its associated challenges remains fragmented. This scoping review examined the applications, reported benefits, and challenges of AI-assisted science lesson planning, with particular attention to its potential contribution to innovative pedagogy in resource-constrained settings. Following the existing framework, as advanced by previous researchers in about 2 decades ago and reported in accordance with PRISMA-ScR guidelines, a systematic search of peer-reviewed and grey literature published between January 2020 and May 2026 identified 40 relevant studies. The synthesis revealed four primary thematic categories of AI application in science lesson planning: (1) generating lesson materials and content, (2) facilitating virtual laboratory experiences, (3) automating planning and documentation, and (4) supporting pedagogical decision-making. Across the reviewed literature, reported benefits included reduced lesson-preparation time, improved access to curriculum-aligned instructional resources, and support for inquiry-based and differentiated teaching approaches. Reported challenges included inaccurate or fabricated AI-generated information, insufficient contextualization of instructional materials, unequal access arising from digital and infrastructural constraints, and variation in teachers' capacity to critically evaluate AI-generated outputs. The findings suggest that AI can support science lesson planning and contribute to innovative pedagogical practices when used as a complementary tool rather than a substitute for teachers' professional judgment. The review indicates that effective and responsible implementation requires reliable digital infrastructure, critical verification of AI-generated content, contextual adaptation of AI outputs, and continuing teacher professional development. These findings highlight both the potential of AI to address resource constraints in science education and the need for context-sensitive approaches to its integration into lesson planning.

Keywords: AI; innovative pedagogy; science education; lesson planning.

Citation: Nsabeyezu, E'., Habakubaho, G., & Mweruli, F. (2026). AI-Assisted science lesson planning for innovative pedagogy in resource-constrained settings: A scoping review. *INES Scientific Journal*, 1(1):28-48
doi: *https://doi.org/10.32677/isj.v1i1.8512*

1. INTRODUCTION

Lesson preparation is one of the most significant components of science teaching and learning (Kraft and Hernández Rodríguez, 2024). It is the stage at which curriculum standards, available materials, students' prior knowledge, and a teacher's own subject-matter and pedagogical content knowledge (PCK) are synthesized into a sequence of classroom activities (Chen and Zhang, 2019). Lesson planning is a critical but resource-intensive component of science teaching, requiring teachers to translate curriculum objectives into appropriate content, learning activities, instructional resources, and assessment strategies. The task can become particularly demanding when teachers have limited access to laboratory facilities, instructional materials, digital resources, or sustained professional support. Although effective lesson planning is essential for organizing classroom instruction and responding to learners' needs, teachers often receive limited structural support for this work after initial teacher training (Hejji Alanazi, 2019).

These persistent demands create a context in which emerging artificial intelligence (AI) tools may offer opportunities to support teachers in developing instructional materials, adapting learning activities, and designing more innovative pedagogical approaches. The burden is amplified in science subjects, where teachers must additionally translate abstract theoretical content like atomic structure, chemical equilibrium, genetic inheritance, energy transformation into concrete, age-appropriate demonstrations, analogies, and where possible, hands-on investigation (Windschitl and Stroupe 2017; Mafugu, Mafugu, and Makwara 2024). In well-resourced education system, this problem manifests primarily as a time and workload problem: teachers report that adapting, searching for, and differentiating instructional materials consumes hours each week that compete with assessment, feedback, and direct teaching time (Karki, 2024). In LMICs and many rural or historically under-funded schools, resource constraints directly affect science lesson planning. Limited laboratory facilities restrict the inclusion of practical activities, while inadequate textbooks constrain access to appropriate content and teaching resources.

Large multigrade classes make differentiated lesson design more demanding, and poor internet connectivity limits access to digital resources and simulations. Shortages of subject-specialist teachers can further constrain the development of scientifically accurate and pedagogically appropriate lessons (Olaifa et al., 2026). Reviews of science teaching in low-resource high schools describe biology, chemistry and physics instruction delivered almost entirely through chalkboard and textbook, with few or no practical activities, even though hands-on experimentation, observation, and iterative inquiry are widely regarded as central to scientific literacy (Zinger et al., 2020; Olaifa et al., 2026). The result is a long-standing equity gap in which the students with the greatest need for engaging, inquiry-rich science instruction are taught by the teachers with the least preparation support (Lupeja and Makwinya, 2021). Since the public release of large language model (LLM)-based chat interfaces in late 2022, generative AI has been adopted across nearly every domain of educational practice, including learning, teaching, assessment, and administration (Ivanova and Terzieva, 2026).

Teachers use generative AI to support science lesson planning by developing lesson materials, examples, quizzes, rubrics, and multimodal resources, while drawing on AI-generated explanations, hints, and iterative feedback to refine and improve their instructional materials (Kim, 2026). For science teachers specifically, generative AI offers a qualitatively new kind of support: rather than retrieving a fixed resource (as a textbook or search engine does), an LLM can generate a tailored analogy, translate a concept into multiple grade levels, produce a step-by-step demonstration script, or draft an inquiry-

based activity on demand, conditioned on a teacher's own prompt describing the topic, grade level, and available materials (Noviello, 2025). This capability is especially relevant in resource-constrained environments, where AI-assisted technologies are increasingly presented as tools that can help transform the educational landscape by increasing accessibility to formats like personalized instruction and distance learning, as well as by helping teachers create efficient study plans and choose suitable learning methodologies (Gupta, 2025). However, a growing body of critical literature warns that teachers' effective use of AI technologies may be hampered by the same issues that are driving AI adoption in resource-constrained educational settings, such as poor digital infrastructure, erratic internet connectivity, a lack of funding, and low levels of digital literacy (Luan et al., 2020). Furthermore, integrating AI might worsen rather than reduce current educational disparities if intentional efforts are not made to improve digital infrastructure, develop instructors' AI competences, localize AI-generated content, and address algorithmic bias (Farooqi et al., 2024).

The body of research on generative AI in education has grown so quickly that a number of systematic and scoping evaluations on related topics have already surfaced: generative AI in K–12 education generally by Sharma et al. (2024), AI-supported teacher professional development by Kusmawan (2023), generative AI in pre-service science teacher education by Spasopoulos (2025), and generative AI personalization in computer science education by Reihanian (2026), among others. These reviews establish that generative AI is reshaping teacher workflows and surface recurring concerns about vision, bias, and overreliance. However, existing reviews and studies appear to give limited focused attention to science lesson preparation as a distinct function of AI-supported teaching. Resource constraints are also often treated as contextual conditions rather than as an explicit analytical lens for examining the potential role of AI in lesson preparation. The present review extends this literature by specifically synthesizing evidence on AI-assisted science lesson planning, examining the reported applications and pedagogical benefits of AI, and analyzing the challenges associated with its use in resource-constrained settings.

First, lesson preparation has distinctive evidentiary requirements that differ from in-class AI use (such as student-facing tutoring) or post-hoc assessment (such as automated grading): the relevant quality criteria concern alignment with curriculum standards, scientific accuracy, pedagogical sequencing, and differentiation, rather than real-time interaction quality (Webb, 2007). Second, science education presents unique challenges that distinguish it from other subject areas. Unlike disciplines such as the humanities or languages, where much of the existing AI-in-education research has been concentrated, science teaching depends heavily on the accurate explanation of scientific concepts and natural phenomena (Choi, 2025). Consequently, the use of AI raises particular concerns about the potential generation of inaccurate, misleading, or oversimplified scientific explanations, which, if incorporated into classroom instruction, could negatively affect students' conceptual understanding (Yannier et al., 2020; Erduran, 2023; Erduran & Levrini, 2024).

Third, focusing on resource-constrained educational settings provides an opportunity to compare how AI is being used across different contexts. Research mostly highlights AI's potential to lessen teachers' workloads, improve instructional efficiency, and personalize learning in high-income, well-resourced educational institutions. On the other hand, research from low-resource and Global South settings frequently sees AI as a way to deal with enduring issues including a lack of trained educators, restricted access to educational resources, and poor educational infrastructure (Muthukrishna et al., 2025). When these two viewpoints are examined inside a common analytical framework, significant contrasts in priorities, presumptions, and implementation issues that have rarely been

examined together are revealed in addition to shared prospects for enhancing teaching and learning.

This scoping review organizes the body of research on AI-supported science lesson preparation by seeing resource scarcity as a cross-cutting analytical component rather than a specialized topic. Arksey and O'Malley (2005) and Levac et al. (2010) advocate the following approach for scoping reviews. Our goal is broad mapping, not effect-size estimation: we want to describe the type, extent, and distribution of this material, specify how AI is thought to help (or hinder) lesson preparation, and identify any gaps that can guide future practice and research. The review is organized around four research questions such as what types of AI tools and applications are reported in the literature on science lesson preparation? what functions does AI serve in the lesson-preparation process and how does this vary between resource-rich and resource-constrained settings? and what risks, limitations, and capacity-building needs does the literature identify?

2. METHODS

This review followed the methodological framework for scoping reviews originally proposed by Arksey and O'Malley (2005) and subsequently refined by Levac et al. (2010), and is reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR; Tricco et al., 2018), a 20-item checklist developed to improve the methodological and reporting quality and transparency of scoping reviews. A scoping rather than effect-size-focused systematic review approach was adopted because the purpose of this review was to map the breadth of AI applications, mechanisms, and reported outcomes in an emerging and heterogeneous body of literature. The approach also enabled the identification of reported risks, contextual factors, and evidence gaps that can inform future primary research, rather than statistically pooling effect sizes from a relatively narrow set of methodologically comparable studies.

2.1 Eligibility criteria

The population comprised studies on school science teachers or studies reporting on teachers' use of AI for science instruction. During screening, studies were considered to address the lesson-preparation dimension when they examined teachers' use of AI to design or develop lesson plans, learning objectives, instructional content, teaching activities, examples, assessments, learning resources, or other materials prepared before classroom instruction. The concept comprised the use of artificial intelligence, including generative AI and large language models, to support science lesson preparation. The context included formal science education settings, with particular attention to resource-constrained environments. Studies were retained when these elements were sufficiently evident in the title, abstract, keywords, or full text (Arksey and O'Malley, 2005). Population: Studies focused exclusively on AI use by students with no teacher-preparation dimension were excluded unless they reported teacher-side lesson-preparation processes. Concept: Use, design, or evaluation of artificial intelligence tools (including but not limited to generative AI/large language models, intelligent tutoring system authoring tools, AI-augmented virtual or simulated laboratories, and AI-based curriculum-recommendation systems) in the preparation, design, adaptation, or planning of science lessons, prior to or independent of classroom delivery. Context: Any institutional or national setting, any science field (physics, chemistry, biology), and any educational level from elementary school to programs for teacher certification.

In keeping with the review's comparative objective, studies from both resource-rich and resource-constrained contexts were eligible. During screening and data charting, contexts were classified based on the availability of key educational resources, including science

laboratories, instructional materials, digital devices, reliable internet connectivity, and subject-specialist teachers, with studies lacking sufficient information classified as "not clearly specified" rather than inferred. Eligible sources included peer-reviewed journal articles, peer-reviewed conference proceedings, and significant grey literature published between January 2020 and May 2026. Grey literature was considered eligible when it provided substantive empirical, evaluative, policy, or practice-based evidence from recognized educational institutions, government agencies, international organizations, research institutes, or professional bodies and contained sufficient information for data extraction; opinion pieces, promotional materials, news articles, brief webpages, and documents lacking adequate evidentiary or methodological information were excluded.

2.2 Information sources and search strategy

A structured search was conducted across five bibliographic databases such as Scopus, Web of Science Core Collection, ERIC, IEEE, supplemented by Google Scholar and the grey-literature sources UNESCO Digital Library and arXiv (relevant to recent preprints on teacher-AI collaboration systems in low-resource schools). The search string combined three concept blocks using Boolean operators: (1) AI terms ("artificial intelligence" OR "generative AI" OR "large language model*" OR "ChatGPT" OR "LLM"); (2) lesson-preparation terms ("lesson plan*" OR "lesson preparation" OR "lesson design" OR "instructional design" OR "curriculum planning"); and (3) science-education terms ("science education" OR "science teach*" OR "chemistry education" OR "biology education" OR "physics education" OR "STEM education"). Database-specific syntax (e.g., MeSH terms in PubMed, controlled vocabulary in ERIC) was adapted accordingly. The reference lists of all included sources, and of closely related prior reviews identified during screening, were hand-searched (citation chaining) to identify additional eligible studies not captured by the database search. The complete flow of records through this process is reported in Figure 1 (Section 3.1), following PRISMA-ScR reporting conventions (Tricco et al., 2018).

2.3 Selection of sources of evidence

All records identified through the search were exported to a reference manager and de-duplicated (n = 214 duplicates removed from 938 identified records). Two reviewers independently screened titles and abstracts against the eligibility criteria using a calibrated screening form. Before the main screening, a pilot round involving a random sample of 50 records was conducted to assess inter-reviewer consistency. Agreement was evaluated by comparing reviewers' inclusion and exclusion decisions, and disagreements were discussed and resolved through consensus, with the eligibility criteria refined where necessary. During the main screening, inter-reviewer agreement was monitored periodically by comparing decisions on a subset of records, while unresolved disagreements were adjudicated through discussion with a third reviewer. This process was used to maintain consistency in applying the eligibility criteria throughout screening. Discrepancies at the title/abstract stage were resolved by retaining the record for full-text review rather than premature exclusion, consistent with scoping-review guidance favoring sensitivity over specificity at early screening stages (Levac et al., 2010). Of the 724 records screened, 612 were excluded at title/abstract stage and 112 reports were sought for full-text retrieval; five could not be retrieved in full text. Full texts of the remaining 107 records were then assessed independently by two reviewers against the full eligibility criteria, with disagreements resolved through discussion and, where necessary, consultation with a third reviewer. Of these, 73 were excluded at full-text review, yielding 34 included sources from database and grey-literature searching. Citation chaining of these 34 sources and of closely related prior

reviews identified 11 additional candidate records, of which 6 met the eligibility criteria, for a final corpus of 40 included sources. The PRISMA-ScR flow of records through identification, screening, eligibility assessment, and inclusion is summarized in Figure 1 (Section 3.1).

2.4 Data charting process and items

A structured data-charting form was developed iteratively: an initial version was piloted on ten included sources by two reviewers working independently, compared, and revised to improve clarity and consistency before being applied to the full corpus, consistent with the iterative charting process recommended by Levac et al. (2010). Extracted items included: citation details; country/region and income classification (using World Bank income-group categories as a proxy indicator of resource context, supplemented by author-reported infrastructural detail where available); science discipline and school level; study design (e.g., quasi-experimental, qualitative case study, design-based research, conceptual/position paper, mixed-methods); AI tool(s) used or proposed; the specific lesson-preparation function(s) addressed; reported outcomes and the evidence underpinning them; reported risks, limitations, or negative findings; and any theoretical framework employed (Thohir et al., 2022; Bebis Jr, 2025).

2.5 Synthesis approach

Data were synthesized narratively in keeping with the descriptive and exploratory orientation of scoping reviews. Findings from the included studies were first extracted and organized according to study characteristics, AI applications, lesson-preparation functions, reported benefits, challenges, and contextual factors. The findings were then compared across studies and grouped into recurring thematic categories based on similarities in the purposes and uses of AI in science lesson preparation. Particular attention was given to similarities and differences between resource-rich and resource-constrained contexts, as well as differences in reported outcomes and implementation challenges. Relationships among AI applications, pedagogical practices, resource conditions, and reported benefits or risks were subsequently examined to identify common patterns, contextual variations, and gaps in the evidence. The resulting themes were iteratively reviewed against the extracted data to ensure that the synthesis reflected the evidence reported in the included studies rather than assumptions made by the authors (Zhu et al., 2022). We employed a sociotechnical synthesis framework, derived inductively from the charted data and modified through discussion among the review team that organizes findings along three interconnected dimensions. (a) Infrastructure conditions (connectivity, hardware, laboratory access, language/localization); (b) pedagogical mechanisms (how AI output is generated, evaluated, and integrated into a lesson artifact); and (c) human-capacity conditions. This sociotechnical framework was developed to interpret how AI-supported science lesson planning is shaped by the interaction of technology, pedagogy, and contextual conditions. It draws conceptually on TPACK, particularly its emphasis on the interrelationship between technological, pedagogical, and content knowledge, but TPACK was not used as a formal coding framework or as a basis for categorizing the included studies. Instead, TPACK served as a conceptual point of reference, while the proposed sociotechnical framework guided the interpretation of how AI applications, teachers' pedagogical practices, and contextual factors such as resource availability, infrastructure, and teacher capacity interact to influence lesson-planning practices and reported outcomes (Mishra & Koehler, 2006). The corpus was characterized using quantitative descriptive statistics (e.g., counts of studies by region, discipline, and research design) and qualitative thematic analysis (Mathis et al., 2024).

2.6 Critical appraisal

Consistent with PRISMA-ScR and JBI guidance, formal risk-of-bias or methodological-quality scoring was not applied as an inclusion filter, since scoping reviews aim to map the full breadth of available evidence regardless of methodological strength (Tricco et al., 2018). However, for each included empirical study we recorded sample size, study duration, and design type to allow readers to weight the strength of specific claims, and we explicitly flag in the narrative synthesis.

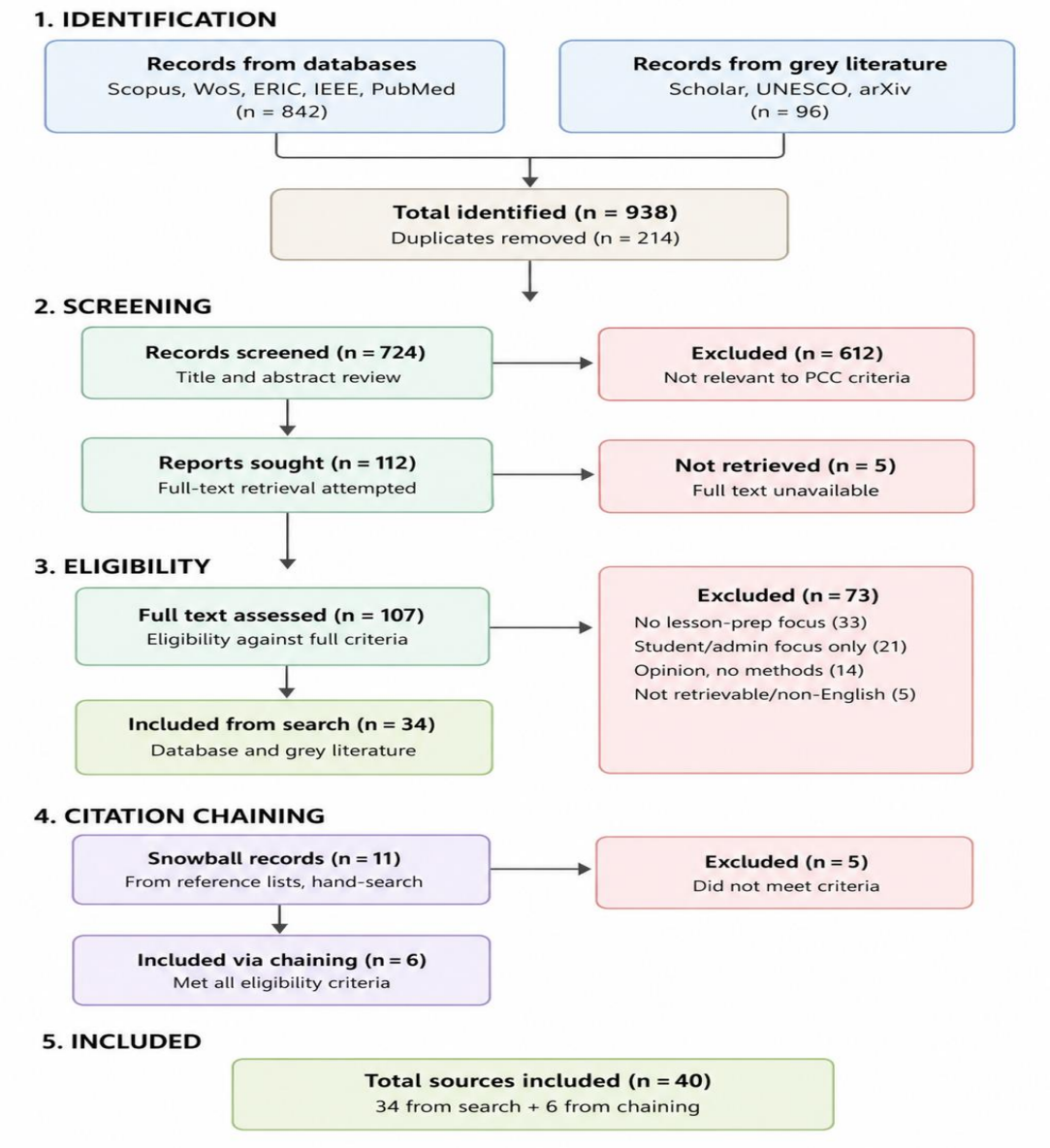


Figure 1. PRISMA Aligned Methodological Parameters

2.7 Characteristics of included sources

The database and grey-literature search, after de-duplication, title/abstract screening, and full-text assessment, yielded a corpus of 40 included sources (Figure 1) characterized by rapid year-on-year growth: less than a tenth of eligible sources were published before the November 2022 release of consumer-facing LLM chat interfaces, with the remainder concentrated in 2023 through early 2026. This trajectory mirrors patterns reported in adjacent reviews of generative AI in K-12 and teacher-education contexts, where the evidence base is described as small but rapidly expanding, with notable concerns about sample size and the underrepresentation of in-service (as opposed to pre-service) teachers in empirical designs.

Geographically, the corpus reflects two overlapping but distinct literatures. The first, more established literature is anchored in North America, Europe, and East Asia and is organized primarily around teacher workload, personalization, and TPACK-based competency development among pre-service teachers (Tokmak et al., 2013). The second, more recent literature is anchored in Sub-Saharan Africa, South Asia, and the Middle East and is organized primarily around infrastructural compensation: virtual laboratories, AI-assisted localization and translation, and teacher-AI collaboration systems designed explicitly for low-resource schools (Varuvel Dennison et al., 2026). Studies addressing chemistry and biology were most numerous, followed by physics and integrated/general science; this distribution likely reflects both the prominence of laboratory-dependent practical work in chemistry and biology curricula and the corresponding salience of resource scarcity in those subjects.

3. RESULTS

3.1 Types of AI tools and their evolution

Throughout the corpus, three major types of AI tools were found. The first and most common category consists of general-purpose conversational generative AI tools, primarily ChatGPT, along with Google Gemini and other LLM-based chat interfaces that teachers directly use through free-text prompting to create assessment items, differentiated materials for multiple grade levels, lesson objectives, and explanatory content (Georgia et al., 2024). Comparative content analyses of lesson plans created by various LLMs (e.g., ChatGPT versus Google Gemini) reveal that output quality and style vary significantly by model, suggesting that conclusions related to one LLM should not be assumed to automatically generalize to others. This is a crucial warning for a rapidly evolving body of evidence that is largely based on a single dominant tool at any given time (Baytak, 2025).

The second category comprises AI-augmented virtual and simulated laboratories, distinguished from traditional virtual laboratories by the presence of an explicit AI function that contributes to the learning or instructional process. For this review, a virtual laboratory was classified as AI-augmented when it incorporated AI-based capabilities such as automated analysis of learner performance or experimental data, adaptive feedback or hints, intelligent task recommendations, personalized learning pathways, or AI-supported interpretation of experimental results. In contrast, virtual laboratories that only provided digital simulations of experiments without an AI-based adaptive, analytical, or decision-support function were classified as traditional virtual laboratories and were not included in this AI-specific category. This distinction was applied during data extraction and

charting based on the functions explicitly reported in each study.

According to Kolil et al. (2025), virtual laboratories (VLs) are digital simulations of traditional laboratories that enable students to interact with virtual representations of real-world objects and instruments via computers or smartphones. VLs are specifically positioned in this literature as a response to the lack of classrooms, libraries, and particularly science laboratories in rural and Global South schools (Shambare & Jita, 2025). AI and machine learning are being used by several VL platforms to assess student behavior and provide customized help. For instance, if a student consistently connects a circuit improperly, the system can offer a focused hint or micro-tutorial instead of a general error message (Sargiotis, 2024; Archita & Saravanan, 2025). From a lesson-preparation perspective, the reviewed evidence suggests that AI-augmented simulations may alter how teachers plan practical science activities by shifting some preparation from developing demonstrations from scratch toward identifying, evaluating, sequencing, and contextualizing existing digital simulations within a lesson. This interpretation is derived from the reported functions of AI-supported virtual laboratory tools and their use in instructional design; it was not directly examined as a planning process across all included studies. Further research is therefore needed to establish how substantially such tools reduce or transform teachers' practical lesson-preparation workload. The third category comprises purpose-built teacher-AI collaboration and curation platforms, distinct from general-purpose chatbots in that they are designed specifically for the lesson-planning workflow, often with curriculum-alignment features built in. One such system, Shiksha Copilot, deployed and studied directly in low-resource government schools across Karnataka, India, generates curriculum-aligned lesson plans in English that curators and teachers then customize (Dennison et al.(2026) Dennison et al. (2026) evaluated this system in a large-scale mixed-methods study involving 1,043 teachers and 23 curators. Their findings indicated that the quality of lesson plans was influenced by the quality of the AI-generated learning objectives. However, this relationship was primarily identified through the qualitative evidence from curators rather than demonstrated through a statistical analysis of the association. Science curators reported that AI-generated learning objectives were often superficial and did not fully capture the breadth of a topic, requiring substantial manual editing before the resulting lesson plans could be used in classroom practice. This finding is crucial because it documents, with field data rather than self-report alone, exactly where in the lesson-preparation pipeline AI assistance currently breaks down.

3.2 Functions of AI in lesson preparation and variation by resource context

Synthesizing across the corpus, four interrelated functions of AI in science lesson preparation were identified. These functions are not mutually exclusive within a single study, and most resource-constrained-context studies report multiple functions operating together.

(a) Generation of explanatory content, analogies, and differentiated materials

The most reported function is the generation of grade-appropriate analogies, explanatory narratives, and differentiated content for a single science topic across multiple developmental levels (Bewersdorff et al., 2025). A representative example documented in the literature is a prompt asking an LLM to provide analogies and stories

illustrating natural selection, adaptation, variation, and speciation tailored separately for third-, sixth-, ninth-, and twelfth-grade students, with simpler language for younger students and greater scientific depth for older students, presented in tabular form and translated into a second language (G. Lee et al., 2026). This single-prompt, multi-output capability is repeatedly cited as a major efficiency gain relative to manually authoring grade-differentiated materials from scratch, and is described as particularly valuable for multigrade and multilingual classrooms common in low-resource settings (Nyaaba et al., 2026).

Evidence on the instructional quality of this generated content is mixed and generally favorable but qualified. A controlled study of ChatGPT-assisted lesson planning in STEAM arts education involving 13 teachers and 6 professors reported that AI-assisted lesson plans scored higher on a structured quality rubric than traditionally authored plans (mean score 21.0 versus 17.0), alongside reported gains in efficiency and support for interdisciplinary and creative approaches (Luo and Tahir, 2025). In a science-specific context, an analysis of 29 pre-service elementary teachers' ChatGPT-integrated science lesson plans, scored against a modified Technology Integration Assessment Rubric (TIAR) adapted to a TPACK framework, found an average score of 3.29 on a 4-point scale (SD = 0.91)—above the midpoint and indicating reasonably high integration competency but with the lowest sub-score (M = 3.00, SD = 0.93) on the dimension assessing whether teachers exploited the AI tool's full functional potential (Lee and Zhai, 2024), suggesting that pre-service teachers tend to under-use the more sophisticated capabilities of generative AI even when their general TPACK-related integration is sound.

(b) Compensation for missing laboratory infrastructure

The second function, concentrated heavily in the resource-scarcity-focused literature, is the use of AI-augmented virtual laboratories and simulations to substitute for absent or inadequate physical laboratory infrastructure during lesson preparation and delivery. This function directly addresses what is repeatedly identified as one of the most consequential structural barriers in under-resourced science education: a shortage of essential infrastructure that severely limits hands-on experimentation and, in turn, limits students' development of core scientific skills such as questioning, hypothesizing, observing, measuring, and handling data (Shambare & Jita, 2025). Virtual laboratories are explicitly framed as a promising alternative for schools lacking physical science facilities, with many programs freely available and able to function without extensive school internet infrastructure, properties that make them comparatively accessible for under-resourced schools relative to other educational technologies (Olaifa et al., 2026).

A quantitative survey of 186 Life Sciences teachers' perspectives on virtual laboratories in rural classrooms in South Africa's Eastern Cape province, grounded in the Technology Acceptance Model, reported that teachers viewed VLs as a genuinely promising response to infrastructural shortages, while also surfacing concerns about device access, training, and the loss of tactile, hands-on skill development that physical laboratory work provides (Shambare et al., 2025). A separate perspective paper focused on bridging the AI divide in high-school STEM education in developing countries similarly frames AI-powered virtual laboratories and simulations as allowing learners to safely and cost-effectively perform experiments that would otherwise be dangerous, expensive, or simply impossible to run in their school's physical facilities, for example, examining cell structures under a virtual microscope or running a genetics cross-breeding simulation entirely on a computer (Olaifa et al., 2026). From a lesson-preparation perspective, the reviewed evidence suggests that AI-supported virtual laboratory resources may broaden how science lessons can be prepared in resource-constrained settings. Rather than relying exclusively on low-cost physical

demonstrations or omitting practical activities because of limited laboratory resources, teachers may identify and evaluate appropriate simulations, sequence them within a lesson, and use AI or general large language models to develop accompanying worksheets, guiding questions, and assessment tasks. This represents an interpretation of the reviewed evidence rather than a directly observed planning process across all included studies. Although both are often utilized by the same teacher during the same lesson cycle, this is a qualitatively distinct type of AI-supported preparation than the content-generation function mentioned above (Dennison et al., 2025).

(c) Automation of routine planning and documentation tasks

The third function concerns the use of AI to support the administrative and documentation aspects of lesson planning, including formatting lesson plans according to required templates, generating planning records, and preparing standards-alignment statements. Evidence from the reviewed studies indicates that these applications have been reported in both resource-rich and resource-constrained educational settings, although their specific use and reported benefits vary across contexts. In resource-rich settings, AI-supported automation may reduce the time teachers spend on routine planning documentation, while in resource-constrained settings, it may help teachers manage documentation requirements where time, materials, and professional support are limited. These findings suggest that the administrative value of AI extends across different resource contexts, although the magnitude and practical significance of this benefit require further comparative investigation.

Field evaluation of Shiksha Copilot, a teacher-AI collaboration system deployed in low-resource schools, found that it eased bureaucratic workload, reduced lesson-planning time, and lowered teaching-related stress, while simultaneously enabling access to more comprehensive, curriculum-aligned plans than teachers reported being able to produce unaided under severe time constraints (Belkhode et al., 2025). This function is reported as a substantial and relatively uncontroversial benefit across both resource-rich and resource-constrained settings, in contrast to the more contested instructional-quality questions raised by the content-generation function.

(d) Scaffolding of pedagogical reasoning

The fourth and most pedagogically significant function is the use of AI not merely to produce lesson content but to scaffold the teacher's own pedagogical reasoning during the planning process prompting consideration of student misconceptions, alternative explanations, or differentiation strategies the teacher might not otherwise have generated. A mixed-methods case study of 19 third-year pre-service science teachers planning lessons on chemical equilibrium in a resource-constrained South African context, framed using the Refined Consensus Model of Pedagogical Content Knowledge (Carlson et al., 2019), examined how AI influenced 'adaptive pedagogical reasoning' across four dimensions: identification and anticipation of students' thinking, adaptivity in lesson design, justification of instructional decisions, and alignment with learning goals and differentiation (Ndlovu et al., 2026). Using Rasch analysis, the study found that AI-supported planning improved three of the four dimensions, particularly justification of instructional decisions and alignment with learning goals, but left adaptivity in lesson design essentially unchanged, suggesting a specific limit on what current one-shot prompting can scaffold (Ndlovu et al., 2026). This and related studies caution that while pre-service teachers are generally receptive to AI integration, how they adapt AI-generated content through genuine pedagogical reasoning as opposed to passively accepting it remains underexplored and appears to be a developmental competency requiring explicit

scaffolding during teacher preparation (Mnguni and Pule, 2025; Morrison-Love & Patrick, 2025; Ndlovu et al., 2026), since student teachers often struggle to systematically apply theoretical pedagogical knowledge when confronted with complex, ill-structured planning problems even without AI in the loop.

A related strand of evidence comes from studies using ChatGPT-supported critique, rather than ChatGPT-supported generation, as the pedagogical mechanism: pre-service teachers' lesson plans are submitted to an LLM for structured critique, and the teachers then revise their own plans in response (Gursoyand Cai, 2025). This approach is reported to effectively support the development of lesson-planning skills, particularly for cognitive-domain objectives, and has been proposed as a foundational framework for integrating AI into teacher-education programs in a way that positions AI as a Socratic interlocutor rather than as a content generator—a meaningfully different instructional design from the more common 'ask the AI to write the lesson plan' pattern documented elsewhere in the corpus (Gursoy and Cai, 2025).

(e) Reported outcomes and the evidence base underpinning them

Across the reviewed empirical studies, seven recurring outcome categories were identified, consistent with thematic patterns reported in adjacent reviews of generative AI in teacher development: improved efficiency/time savings, gains in teachers' technology-pedagogy integration knowledge (TPACK), increased willingness to attempt higher-order and inquiry-based instructional strategies, measurable (though not always statistically significant) improvements in lesson-plan quality as scored by structured rubrics, gains in teachers' self-reported confidence and self-efficacy, more positive attitudes toward and intentions to continue using AI tools, and growth in general AI literacy. Reported intervention durations in the controlled and quasi-experimental studies were typically short, commonly ranging from a single session to several weeks, which limits confidence in the durability of any observed gains.

Two specific studies were selected to illustrate both the promise and the limitations of the current evidence base because they provide contrasting examples of AI-supported science lesson planning across different educational contexts. They were chosen purposively for their clear reporting of the relevant lesson-planning function and their ability to illustrate differences in the reported benefits and challenges. These studies are therefore illustrative rather than representative of all 40 included studies; their findings should be interpreted alongside the broader thematic synthesis presented in this review. First, a quasi-experimental study evaluating ChatGPT-generated lesson plans for improving primary-school students' mathematics achievement (a closely analogous design to several science-specific studies), involving 39 third-grade students across an experimental and a control group, found large within-group gains for both AI-assisted and teacher-developed plans (Cohen's d ≈ 1.25–1.27) but no statistically significant difference between the two groups' final scores, suggesting that demonstrated improvements in lesson-plan quality on a rubric do not automatically translate into measurable, statistically robust student-learning gains, and that this causal chain from AI-assisted planning, to lesson-plan quality, to classroom enactment, to student outcomes is rarely traced in full within a single study (Karaman and Göksu, 2025). Second, in a study explicitly addressing disadvantaged students, generative AI used for personalized feedback and lesson planning was associated with improved grades and a narrowing of the attainment gap between disadvantaged (pupil-premium and special-educational-needs) students and their peers, while also reducing the marking and feedback time burden on teachers serving these students (Pargmann et al., 2025).

Notably, some studies in the corpus independently verified the scientific accuracy of AI-generated lesson content against expert or

curriculum-standard benchmarks; most relied on structured rubrics assessing pedagogical and technology-integration dimensions (e.g., TPACK-aligned rubrics) rather than subject-matter fact-checking (Lee and Zhai, 2024; Luo and Tahir, 2025).

3.3 Risks, limitations, and capacity-building needs

The literature converges on a consistent, if not yet fully resolved, set of risks associated with AI-supported science lesson preparation, which we organize into four categories: epistemic risks, equity risks, professional-practice risks, and systemic/infrastructural risks.

Epistemic risks: hallucination and scientific inaccuracy

A major limitation of generative artificial intelligence is hallucination, referring to the generation of plausible-sounding but factually incorrect, unsupported, or fabricated information. This limitation is particularly important in educational contexts because fluent and coherent AI-generated content may appear authoritative despite containing inaccuracies. Large language models generate text by estimating the probability of successive tokens based on patterns and representations learned during training; this generative mechanism is optimized for producing coherent language rather than guaranteeing the factual accuracy of every output. Consequently, an AI-generated response may be linguistically convincing without being factually reliable. Empirical research has demonstrated that such factuality failures can occur even when models have relevant knowledge, highlighting the need for systematic verification of AI-generated information before it is incorporated into science lessons or instructional materials (Guan et al., 2024; Wang et al., 2024). They do not check if what they are saying is true. When these systems do not really know something they will still give an answer of saying they are not sure. This can make it hard for someone who iss not an expert, like a teacher who is busy to tell if the information is wrong. This is especially true for subjects that are changing quickly. This is a deal, for science because if a teacher uses wrong information in a lesson the students might learn things that are not true. It can be hard to fix this on. The research says that teachers should be able to check the information created by artificial intelligence systems to find any mistakes or biases. However, the research also says that teachers are not usually taught how to do this when they are training to be teachers. Just being comfortable using intelligence tools is not enough to know how to check the information they produce.

Equity risks: the prompting-skill and access gradient

The quality of AI-generated instructional content is largely dependent on a teacher's ability to prompt effectively, a skill that is unevenly distributed and only marginally correlated with teaching experience or subject knowledge (Lee & Zhai, 2024). This is a second risk that is less often discussed but conceptually significant. Teachers with more expertise in AI may get a lot more instructional value from the same AI tool than their less experienced colleagues working with comparable pupils since output quality improves with specific suggestions and regular adjustment (Fitria, 2023). If AI literacy instruction is not distributed fairly, there is a chance that achievement and resource gaps between schools or classes would worsen rather than improve (Dangol et al., 2025; Olaifa et al., 2026).

Professional-practice risks: superficial objectives and reduced productive struggle in planning

AI-generated learning objectives that are superficial and do not fully cover a subject, requiring science curators to make substantial revisions before the lesson plan is operational (Dennison et al., 2025). This finding is important because it changes the "AI output needs verification" problem from a general, theoretical caution to a documented, topic-specific trend: Even though surface-level content production (analogies, explanations, activity suggestions) frequently receives positive evaluations, the learning goal, the central component of a lesson plan that guides sequencing, activities, and evaluation is precisely where current systems are found to fall short. The constant availability of AI as a planning assistant may promote a type of professional cognitive offloading, decreasing opportunities for the fruitful struggle through which teachers historically developed their own pedagogical content knowledge and contextual judgment (Dangol et al., 2025; Springer Nature, 2024). This is a related concern that has been raised in the larger generative-AI-in-education literature and is plausibly applicable to teacher lesson preparation itself.

Systemic and infrastructural risks

Limited and unreliable internet connectivity, financial constraints on licensing or subscription costs for more capable AI tools, inadequate teacher training infrastructure, and the predominance of English-language and high-income-context training data underlying most widely used LLMs, which can produce content poorly localized to local curricula, languages, and cultural reference points (Farooqi et al., 2024; Olaifa et al., 2026). Without deliberate investment in digital infrastructure, content localization, and bias mitigation, the literature cautions that AI adoption in education may track and reinforce existing global disparities in technology adoption rather than closing them, echoing concerns raised about AI adoption gaps between low-income and high-income countries more broadly across sectors beyond education alone (Springer Nature, 2024).

Table1 Summary of AI Functions, Representative Mechanisms, and Associated Risks in Science Lesson Preparation

Function	Representative Mechanism	Primary Reported Risk
Content & analogy generation	Multi-grade differentiated explanations from a single prompt (e.g., evolution analogies for grades	Hallucination; scientific inaccuracy requiring expert verification
Infrastructure compensation	AI-augmented virtual laboratories/simulations replacing absent physical labs	Loss of tactile/hands-on skill development; device/connectivity access gaps
Administrative automation	Auto-formatted, curriculum-aligned plans and records via teacher-AI platforms	Superficial learning objectives requiring manual rewriting
Pedagogical reasoning scaffolding	AI-supported critique of teacher-authored plans; adaptive pedagogical reasoning prompts	Underdeveloped critical-adaptation skill among novices; risk of passive acceptance

4. DISCUSSION

4.1 AI as a contingent amplifier, not a substitute, for pedagogical content knowledge

The most consistent pattern in this corpus is that AI's impact on science lesson preparation is contingent rather than uniform: the same tool produces meaningfully different outcomes depending on the teacher's pedagogical content knowledge, verification practices, and institutional context. Pre-service teachers with stronger TPACK-related integration competency are reported to produce higher-quality AI-assisted plans and to exploit a greater share of an AI tool's functional capability (Lee & Zhai, 2024); teacher-AI collaboration platforms produce usable lesson plans only after subject-specialist curators substantially rework AI-generated learning objectives (Dennison et al., 2025); and structured AI-critique protocols appear to build pedagogical reasoning skills (Cai et al., 2025; Ndlovu et al., 2026). Taken together, these findings support reframing the central question of this literature away from 'does AI improve science lesson preparation?' and toward 'under what pedagogical and institutional conditions does AI improve science lesson preparation, and for whom?' This reframing aligns with, and extends to the lesson-preparation context specifically, the broader technology-integration literature's longstanding insight formalized in the TPACK framework—that technology's instructional value is inseparable from the pedagogical and content knowledge brought to bear on it (Mishra & Koehler, 2006).

4.2 Resource scarcity: genuine compensatory potential alongside a real risk of widening inequity

This review's explicit focus on resource scarcity surfaces a genuine point of tension within the literature rather than a single, settled conclusion. On one hand, the evidence on AI-augmented virtual laboratories and teacher-AI collaboration platforms supports a credible compensatory narrative: schools with no physical laboratory, intermittent electricity, or a single generalist teacher covering multiple science subjects and grades can, in principle, access simulated practical work and curriculum-aligned planning support that would otherwise be entirely unavailable to them (Dennison et al., 2025; Shambare et al., 2025), and field deployments report concrete reductions in the documentation burden that disproportionately falls on teachers in under-resourced, high-administrative-load systems (Dennison et al., 2025). On the other hand, the same body of literature consistently flags conditions that could just as easily reproduce or deepen existing disparities: dependence on the prompting skill and AI literacy of an already-overstretched teacher (Lee and Zhai, 2024); dependence on connectivity and devices that remain unevenly distributed even within low-resource regions (Shambare et al., 2025); and a structural tilt in most widely used LLMs toward English-language, high-income-context training data that can leave AI-generated content poorly localized to local curricula, languages, and cultural reference points unless deliberately corrected for (Nyaaba et al., 2025; Olaifa et al., 2026). We see this tension not as a contradiction in evidence, but as a description of a genuine fork in possible futures, the result of which is determined by policy and investment decisions made mainly outside of the AI tool itself. The literature's repeated calls for partnering AI initiatives with local educators and institutions, investing in digital infrastructure, and purposefully localizing content (Nyaaba et al., 2025; Olaifa et al., 2026) are best understood as the conditions under which the compensatory potential of AI-supported lesson preparation is more likely to be realized than its inequity-widening potential (Vladimirov & Harizanova, 2026).

4.3 A practitioner-facing framework for critically appraising AI-generated science lesson content

By combining the risks associated with the established mechanisms of effective utilization, we propose a four-step evaluation procedure for science teachers and teacher educators to use when incorporating AI output into a lesson, as shown in Table 2. The sequence is designed to address a specific, documented failure point in this literature: shallow or inaccurate learning objectives and explanatory material, before that content influences activity design and assessment. The four-step evaluation methodology highlights that effective preparation of scientific lectures with AI aid is dependent on instructors' rigorous professional appraisal during the lesson planning stages, rather than blindly accepting AI-generated information. The approach begins with validation, which recognizes that generative AI may produce scientifically incorrect or fraudulent information that must be compared to curriculum guidelines and proven scientific facts. It then emphasizes the need to improve learning objectives such that AI-created lesson plans progress from surface-level content to major conceptual comprehension. The concept highlights the importance of tailoring AI outputs so that examples, terminology, and instructional materials are relevant to learners' cultural, linguistic, and resource contexts, particularly in situations with few resources where generic AI replies may be ineffective. Finally, the adapt and justify step emphasizes the importance of educators properly articulating their pedagogical decisions about the appropriateness of AI-generated recommendations for their specific students, rather than accepting them without question. This methodology demonstrates that the usefulness of AI-assisted scientific lesson planning is dependent on instructors' ability to critically examine, contextualize, and pedagogically validate AI-generated material before deploying it in the classroom, rather than the AI's inherent capabilities.

Table 2 Proposed four-step Appraisal Sequence for AI-Assisted Science Lesson Preparation

Step	Guiding question	Rationale (Linked finding)
1. Verify	Is the science content factually accurate against curriculum principles and current disciplinary harmony?	Hallucination is well recognized and not dependably self-flagged by the model
2. Deepen	Does the instructional objective seizure the full conceptual breadth of the topic, or only a shallow slice?	Field confirmation shows AI-generated objectives are often shallow and require knowledgeable rewriting
3. Localize	Are examples, language, and materials suitable to students' context, resource level, and home linguistic?	Generic/high-income-context output risks poor fit in localized or low-resource locations
4. Adapt & justify	Has the teacher made a clear instructional judgment about why this AI-suggested approach fits these students, rather than compliant it as-is?	Adaptive pedagogical reasoning, not passive acceptance, predicts excellence outcomes

4.4 Implications of the study

The findings of this review have implications for teacher education and professional development. While current initiatives place a greater emphasis on teachers' ability to use AI tools and incorporate them into teaching practices through frameworks such as TPACK, evidence suggests that a more urgent requirement is the development of output-validation literacy, or the ability to critically assess

the scientific accuracy, pedagogical standards, logical consistency, and potential biases in AI-generated materials prior to classroom implementation. For science educators whose talents are centered in judging scientific concepts, teacher training programs should transition from just teaching instructors how to develop lesson plans with AI to incorporating organized opportunities to criticize, verify, improve, and justify instructional resources created by AI. Incorporating these critical evaluation abilities into pre-service training and continuous professional development would strengthen teachers' ability to use AI responsibly while maintaining the integrity and accuracy of science instruction. The review identifies several policy-relevant considerations supported by the synthesized evidence, particularly the need for reliable digital infrastructure, teacher professional development, access to appropriate instructional resources, and mechanisms for verifying the scientific accuracy and contextual relevance of AI-generated materials. These findings suggest that AI may provide additional support for access to instructional resources in resource-constrained settings; however, the review does not establish that AI directly reduces educational inequalities. Building on these evidence-based considerations, the authors propose that education systems develop context-sensitive AI guidance that addresses infrastructure, teacher capacity, responsible use, content verification, data protection, and equitable access. These broader policy recommendations represent the authors' interpretation of the evidence and should be further evaluated through empirical research.

Instead of focusing solely on pushing AI adoption, authorities should prioritize creating the circumstances for its successful and equitable usage. This necessitates ongoing investment in dependable internet access, digital frameworks, and shared access to AI technologies, as well as the creation of locally relevant, curriculum-aligned AI materials through collaboration among curriculum designers, subject specialists, and educators. Simultaneously, efforts in professional development should prioritize improving teachers' abilities to critically review and change AI-generated content rather than simply teaching them how to use AI technologies. An integrated approach to AI use can strengthen its pedagogical value while mitigating risks associated with inaccurate content, inadequate contextualization, and inequitable instructional practices.

5. CONCLUSION

This scoping review mapped the reported applications, benefits, and challenges of artificial intelligence (AI) in science lesson preparation, with particular attention to resource-constrained educational settings. Across the reviewed literature, four complementary functions emerged: generating instructional materials and scientific content, supporting virtual laboratory experiences, automating planning and documentation, and supporting pedagogical decision-making. The evidence suggests that AI can assist teachers with several aspects of lesson preparation and may help address some resource-related constraints, including limited access to instructional materials, laboratory resources, and teachers' planning time. However, the reviewed evidence does not establish that AI independently improves science teaching quality or learning outcomes. The findings further indicate that the educational value of AI depends substantially on teachers' ability to evaluate, contextualize, adapt, and validate AI-generated outputs.

Reported concerns regarding factual inaccuracies, inadequate contextualization, unequal access, and variation in teachers' AI-related competencies reinforce the importance of maintaining teachers' professional judgment in AI-assisted lesson preparation. In this review, output-validation literacy is proposed as a synthesis-derived concept referring to teachers' capacity to critically examine AI-generated

instructional content for scientific accuracy, pedagogical appropriateness, contextual relevance, and potential bias before classroom use. This concept should be regarded as a proposition emerging from the synthesis rather than an empirically established competency, and it warrants validation through future primary research. Future research should move beyond examining the characteristics or quality of AI-generated lesson plans to investigate the complete pathway from AI-assisted planning to classroom implementation and student learning. Comparative studies should also examine how AI-supported lesson preparation operates across resource-rich and resource-constrained settings, including differences in infrastructure, teacher capacity, access, and contextual adaptation. Such evidence could inform context-sensitive professional development, institutional guidance, and policies for the responsible integration of AI into science education.

REFERENCES

Archita, S., & Saravanan, S. (2025). Enhancing Student Learning with a Customizable Virtual Assistant. *2025 3rd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, 1–7.

Bebis Jr, M. (2025). Technological, Pedagogical, and Content Knowledge (TPACK) of Pre-Service Teachers and their Preparedness in Teaching Science: Basis for Intervention Plan. *Aloysian Interdisciplinary Journal of Social Sciences, Education, and Allied Fields*, *1*(9), 602–625.

Belkhode, P. N., Awatade, S. M., Prakash, C., Shelare, S. D., Marghade, D., Gajghate, S. S., Noor, M. M., & Dennison, M. S. (2025). An integrated AI-driven framework for maximizing the efficiency of heterostructured nanomaterials in photocatalytic hydrogen production. *Scientific Reports*, *15*(1), 24936.

Bewersdorff, A., Nerdel, C., & Zhai, X. (2025). How AI literacy correlates with affective, behavioral, cognitive and contextual variables: A systematic review. *Computers and Education: Artificial Intelligence*, 100493.

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., & Friedrichsen, P. (2019). The refined consensus model of pedagogical content knowledge in science education. In *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–94). Springer.

Chen, S., & Zhang, B. (2019). Improving prospective teachers' lesson planning knowledge and skills through lesson study. In *Theory and practice of lesson study in mathematics: An international perspective* (pp. 549–575). Springer.

Choi, Y. (2025). Exploring the scientific validity of ChatGPT's responses in elementary science for sustainable education. *Sustainability*, *17*(7), 2962.

Erduran, S. (2023). AI is transforming how science is done. Science education must reflect this change. In *Science* (Vol. 382, Issue 6677, p. eadm9788). American Association for the Advancement of Science.

Farooqi, M. T. K., Amanat, I., & Awan, S. M. (2024). Ethical considerations and challenges in the integration of artificial intelligence in education: A systematic review. *Journal of Excellence in Management Sciences*, *3*(4), 35–50.

Fitria, T. N. (2023). The use of artificial intelligence in education (AIED): Can AI replace the

teacher's role? *Epigram*, *20*(2), 165–187.

Gupta, S. (2025). AI agents collaboration under resource constraints: Practical implementations. *International Journal of Artificial Intelligence Research and Development*, *3*(1), 51–63.

Gursoy, D., & Cai, R. (2025). Artificial intelligence: an overview of research trends and future directions. *International Journal of Contemporary Hospitality Management*, *37*(1), 1–17.

Hejji Alanazi, M. (2019). A study of the pre-service trainee teachers problems in designing lesson plans. *Arab World English Journal (AWEJ) Volume*, *10*(1), 2–8.

Ivanova, T., & Terzieva, V. (2026). Large Language Models in Intelligent Education Systems: New Educational Perspectives—A Systematic Review. *Information*, *17*(5), 433.

Karaman, M. R., & Göksu, İ. (2025). Exploring the Relationship Between Teachers' Artificial Intelligence Awareness and Computational Thinking Skills: Differences Across Sociodemographic and Professional Variables. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, *21*(66), 4394–4431.

Karki, H., Jamtsho, S., & Basnet, M. B. (2024). Block Time Lesson Planning in Managing Teachers' Workload. *Journal of the International Society for Teacher Education*, *28*(1), 9–25.

Kim, W. J. (2026). Teachers ' Use of Generative Artificial Intelligence for Designing Science Lessons in Support of Environmental Science Agency. *Research in Science Education*, *56*(2), 203–222.

Kraft, T., & Hernández Rodríguez, O. (2024). Lesson study in science teacher preparation programs: a literature review. *International Journal for Lesson & Learning Studies*, *13*(2), 116–129.

Kusmawan, U. (2023). Redefining Teacher Training : The Promise of AI-Supported Teaching Practices. *Journal of Advances in Education and Philosophy*, *2665*, 332–335. https://doi.org/10.36348/jaep.2023.v07i09.001

Lee, G.-G., & Zhai, X. (2024). Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. *IEEE Transactions on Learning Technologies*, *17*(1), 1643–1660.

Lee, G., Yun, M., Zhai, X., & Crippen, K. (2026). Artificial intelligence in science education research: Current states and challenges. *Journal of Science Education and Technology*, *35*(1), 110–127.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology*, *11*(1), 580820.

Luo, Z., & Tahir, R. (2025). ChatGPT-Assisted Lesson Planning for Children's STEAM Arts Education: An Experimental Study on Benefits, Challenges, Methods, and a Prompt Framework. *International Journal of Artificial Intelligence in Education*, *5*(1), 1–50.

Lupeja, T. L., & Makwinya, N. M. (2021). The constraints of linking instructional methods to learners' real-life settings in the context of resources scarcity. In *The constraints of linking instructional methods to learners' real-life settings in the context of resources scarcity* (pp. 1–24). AkiNiki Publications New Delhi, India.

Mafugu, T., Mafugu, F., & Makwara, C. (2024). Exploring the challenges faced by preservice teachers in executing science lessons during work integrated learning in rural school settings. *E-Journal of Humanities, Arts and Social Sciences*, *5*(13), 1–15.

Mnguni, H., & Pule, K. F. (2025). Impact of Digital Transformation on the Livelihoods of Subsistence Farmers in South Africa: Insights from Emerging Markets. *IAHRW International Journal of Social Sciences Review*, *13*(6), 1247–1256.

Muthukrishna, M., Dai, J., Panizo Madrid, D., Sabherwal, R., Vanoppen, K., & Yao, H. (2025). AI can revolutionise education but technology is not enough: human development meets cultural evolution. *Journal of Human Development and Capabilities*, *26*(3), 482–492.

Ndlovu, B. P., Khoza, H. C., & Sibanda, D. (2026). Pre-service science teachers' adaptive pedagogical reasoning in AI-supported lesson plans: a case with chemical equilibrium. *Frontiers in Education*, *11*, 1811975.

Noviello, Y. (2025). Designing and Evaluating AI-Generated Multimodal Analogy-Based Explanations. *International Conference on Artificial Intelligence in Education*, 446–451.

Nyaaba, M., Kyeremeh, P., Nabang, M., Akanzire, B. N., Titty, C. A., Acquah, S., Asare, K., & Kudaya, J. E. (2026). Conversational AI agent for teacher education in Ghana. *Discover Education*, *3*(1), 1–12.

Olaifa, K. W., Elshikh, M., Balgaliyeva, A., Sagindykova, E., Inegbedion, E., & Nussipzhanova, G. (2026). Bridging the AI divide in high school STEM education in developing countries to advance equity and innovation. *Discover Education*, *5*(1), 415.

Pargmann, J., Berding, F., Rebmann, K., & Riebenbauer, E. (2025). How AI feedback supports lesson planning in vocational teacher education: a longitudinal intervention study using an analytical AI platform. *Empirical Research in Vocational Education and Training*, *17*(1), 26.

Reihanian, I. (2026). From Pilots to Practices : A Scoping Review of GenAI-Enabled Personalization in Computer Science Education. *AI*, *21*(2), 1–23.

Sargiotis, D. (2024). Integrating ai and big data in virtual infrastructures: Transforming educational landscapes for the future. *Available at SSRN 4789850*.

Shambare, B., & Jita, T. (2025). From desire to understanding: unveiling science teachers' perspectives on virtual laboratories in rural classrooms. *Cogent Education*, *12*(1), 2529079.

Sharma, D. M., Ramana, K. V., Jothilakshmi, R., Verma, R., Maheswari, B. U., & Boopathi, S. (2024). Integrating generative AI into K-12 curriculums and pedagogies in India: Opportunities and challenges. *Facilitating Global Collaboration and Knowledge Sharing in Higher Education with Generative AI*, *23*(1), 133–161.

Spasopoulos, T. (2025). Generative AI in Pre-Service Science Teacher Education : A Systematic Review Literature Review Science Education. *Advances in Mobile Learning Educational Research*, *5*(2), 1501–1523. https://doi.org/10.25082/AMLER.2025.02.007

Thohir, M. A., Jumadi, J., & Warsono, W. (2022). Technological pedagogical content knowledge (TPACK) of pre-service science teachers: A Delphi study. *Journal of Research on Technology in Education*, *54*(1), 127–142.

Tokmak, H. S., Yelken, T. Y., & Konokman, G. Y. (2013). Pre-service teachers' perceptions on development of their IMD competencies through TPACK-based activities. *Journal of Educational Technology & Society*, *16*(2), 243–256.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., & Weeks, L. (2018). PRISMA extension for scoping reviews

(PRISMA-ScR): checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473.

Varuvel Dennison, D., Ahtisham, B., Chourasia, K., Arora, N., Singh, R., Kizilcec, R. F., Nambi, A., Ganu, T., & Vashistha, A. (2026). Shiksha Copilot: Teacher-AI Collaboration for Curating and Customizing Lesson Plans in Low-Resource Schools. *Proceedings of the ACM on Human-Computer Interaction*, *10*(2), 1–47.

Vladimirov, Z. D., & Harizanova, O. B. (2026). Factors influencing the use of AI in SMEs: the importance of digital infrastructure. *European Journal of Management Studies*, *3*(1), 1–17.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, *20*(1), 7–25.

Windschitl, M. A., & Stroupe, D. (2017). The three-story challenge: Implications of the next generation science standards for teacher preparation. *Journal of Teacher Education*, *68*(3), 251–261.

Yannier, N., Hudson, S. E., & Koedinger, K. R. (2020). Active learning is about more than hands-on: A mixed-reality AI system to support STEM education. *International Journal of Artificial Intelligence in Education*, *30*(1), 74–96.

Zinger, D., Sandholtz, J. H., & Ringstaff, C. (2020). Teaching science in rural elementary schools: Affordances and constraints in the age of NGSS. *Rural Educator*, *41*(2), 14–30.