Creative Education, 2025, 16(9), 1364-1401
https://www.scirp.org/journal/ce
ISSN Online: 2151-4771
ISSN Print: 2151-4755

From Support Tool to Learning Partner: A Systematic Review of GenAI Integration in University Science Labs

Michail Kalogiannakis, Nikolaos Papakonstantinou, Dimitrios Sotiropoulos

Department of Special Education, University of Thessaly, Volos, Greece
Email: mkalogian@uth.gr, nipapakonst@uth.gr, dsotiropoulos@uth.gr

How to cite this paper: Kalogiannakis, M., Papakonstantinou, N., & Sotiropoulos, D. (2025). From Support Tool to Learning Partner: A Systematic Review of GenAI Integration in University Science Labs. *Creative Education, 16,* 1364-1401.
https://doi.org/10.4236/ce.2025.169083

Received: July 24, 2025
Accepted: September 8, 2025
Published: September 11, 2025

Abstract

This Systematic Literature Review (SLR) investigates the integration of generative artificial intelligence (GenAI) tools, such as ChatGPT, into science education (SE) and STEM education at the higher education level, with a particular focus on experimental and laboratory-based learning environments. Adhering to the PRISMA protocol, 164 initial records were screened, and 20 empirical studies published between 2023 and 2025 were selected based on strict inclusion criteria: relevance to generative AI, application in SE or related STEM fields, involvement of students or pre-service teachers, integration into hands-on instruction, and empirical validation. This selection process ensured both conceptual relevance and methodological rigor. The thematic analysis was structured around five key domains: cognitive benefits, pedagogical practices, student attitudes, teacher perceptions, and broader interpretive convergences. The findings indicate that GenAI supports conceptual understanding, problem-solving, and the development of metacognitive strategies when implemented within well-designed pedagogical frameworks. While students generally held positive attitudes toward the use of GenAI, they also expressed concerns about the reliability and critical interpretation of AI-generated content. Educators recognized GenAI's educational potential but underscored the need for guidance, scaffolding, and professional development. The review highlights two major interpretive tensions: the simultaneous presence of acceptance and uncertainty among learners, and the dual role of GenAI as either a metacognitive facilitator or a superficial answer generator, depending on its use context. Despite methodological limitations related to sample size and study duration, the selected studies offer valuable insights into the responsible and effective use of GenAI in science education, informing both future research and instructional design. It also emphasizes the need to align GenAI integration with evidence-based teaching strategies to create meaningful and equitable learning experiences in science education and in broader STEM contexts.

Keywords

ChatGPT, Generative Artificial Intelligence, Higher Education, Laboratory-Based Instruction, Science Education, STEM Education

1. Introduction

In recent years, the integration of digital technology into education has been a significant factor in the qualitative improvement of education, aiming to equip students with the necessary cognitive and metacognitive tools to meet the challenges of the digital age (OECD, 2023). The use of digital technologies enhances teacher effectiveness (Harris et al., 2009; Öztop, 2023; Timotheou et al., 2023), while their use by students is associated with increased academic performance and the development of research and collaboration skills (Delgado et al., 2015). Within this broader digital transformation, particular emphasis has recently been placed on the use of artificial intelligence (AI), specifically generative artificial intelligence (GenAI) tools, which are increasingly penetrating educational practice.

Generative AI tools such as ChatGPT have recently emerged as a focal point in educational research, regardless of subject area. These applications are already being used to enhance learning, generate content, provide feedback, and support both students and instructors across different levels and contexts (Collie & Martin, 2025; Fan et al, 2024). At the higher education level, the adoption of GenAI has increased, with students and instructors using it for natural language interaction, support in writing and conceptual understanding, and enhancement of self-regulated learning (Kotsis, 2024). Particularly noteworthy is the application of GenAI in undergraduate STEM courses, where it has been successfully used to guide students in project-based environments and support the development of conceptual understanding, for example, in thermodynamics (Wu et al., 2025). Moreover, examples from educational practice suggest that GenAI tools can contribute to student reflection and deeper engagement with scientific content, while also enhancing instructional guidance and the role of educators (Lee & Zhai, 2024).

However, the focus on highly abstract and conceptually demanding subjects, such as Science Education, remains limited. Despite the presence of encouraging examples, the research literature has yet to adequately map how GenAI is utilized in science teaching, particularly when it occurs within experimental or laboratory-based approaches (Lee & Zhai, 2024). This (SLR) aims to address this gap by focusing on empirical studies that examine the attitudes and experiences of university students and educators regarding the use of GenAI as a supportive tool in Science and STEM education.

2. Literature Review

2.1. Artificial Intelligence in Education

Artificial Intelligence (AI) holds significant potential; however, its integration into educational settings presents various challenges that require pedagogically grounded strategies, along with clear ethical and institutional guidance (Williamson & Eynon, 2020). Moreover, AI systems can provide valuable support to students with special educational needs by leveraging tools such as automatic transcription, speech recognition, and translation (Kent & du Boulay, 2022). In addition, by upgrading and enriching the learning environment, AI can enhance students' autonomy and stimulate their creativity (Huang et al., 2021). The benefits are not limited to students, as educators can also benefit. AI's analytical capabilities enable it to monitor student progress, detect behavioural patterns and signs of disengagement, while also supporting the design of effective instructional strategies and targeted feedback. AI does not replace the pedagogical role of the teacher, but rather enhances their ability to meet students' needs (Montenegro-Rueda et al., 2023).

2.2. Generative AI and ChatGPT: Features and Capabilities

GenAI is part of the broader family of Artificial Intelligence. It includes systems capable of generating new content such as text, images, or audio through training on large datasets. Tools such as OpenAI's ChatGPT, Anthropic's Claude, or Google's Gemini enable users to interact through natural language and receive information or guidance in real-time (UNESCO, 2023; European Union, 2024).

ChatGPT-4, in particular, is a large language model based on the GPT-4 architecture. It has been trained on a broad range of data in order to process and generate natural language, performing tasks such as text generation, concept explanation, and question answering (OpenAI, 2024). Although it operates on a predictive basis, the ability to anticipate the next appropriate word during text generation rather than intelligence in the strict sense methods such as Reinforcement Learning from Human Feedback (RLHF) have improved its accuracy and adaptability (Bubeck et al., 2023).

Several recent studies suggest that the use of GenAI tools is associated with enhanced learning engagement, the cultivation of creative thinking, and the development of collaborative skills, particularly when these tools are integrated into learning environments that promote interaction and reflection (Candrasari et al., 2024; Zhou & Li, 2023). Furthermore, applications such as ChatGPT appear to support conceptual understanding, argument construction, and the development of written expression through dialogic interaction, while simultaneously offering access to helpful information directly and functionally (Bubeck et al., 2023). Numerous studies have reported that GenAI can provide real-time guidance, personalized to user needs, and varied feedback, which contributes to sustained interest and improved academic performance among students (Ai et al., 2024; Ugras et al., 2024).

At the same time, from the educators' perspective, GenAI is recognized as a support tool in instructional design and assessment, enabling the identification of weaknesses, the suggestion of alternative formulations, and the provision of corrective feedback (Ding et al., 2024; Nguyen et al., 2024). However, it is emphasized that this technology should not be integrated uncritically, but instead utilized within a pedagogically sound and ethically supervised framework, so that it enhances the teacher's capabilities rather than replacing them (Petousi & Sifaki, 2020; Montenegro-Rueda et al., 2023; Collie & Martin, 2025).

2.3. Risks, Limitations, and Ethical Dilemmas of GenAI

Despite its impressive capabilities, the appropriate use of GenAI remains a pressing concern for its sustainable integration into educational practice. Educators and researchers alike have raised substantial reservations about its reliability and pedagogical value. As several studies point out, GenAI often produces responses that lack logical consistency, contain factual inaccuracies or misleading content, and may reflect embedded biases (Rahman & Watanobe, 2023). Furthermore, its limited ability to process visual and numerical inputs poses challenges in domains that require high-level cognitive engagement, such as science and mathematics, where deficiencies in mathematical reasoning have also been observed. The inherently mechanistic nature of language generation driven by probabilistic modelling rather than genuine comprehension (OpenAI, 2024) intensifies doubts about the epistemic value of its outputs (Collie & Martin, 2025). Within this context, many educators voice concern over the potential of GenAI to displace human guidance, particularly when students engage with such tools as authoritative sources without applying critical judgment (Ausat et al., 2023; Yoon et al., 2024).

Additionally, there are growing concerns that students may become overly dependent on GenAI tools, which could hinder their creativity, analytical thinking, and ability to solve problems on their own (Chinonso et al., 2023; Valeri et al., 2025). In some cases, learners tend to accept responses from ChatGPT without question, showing little engagement in metacognitive thinking or personal interpretation. This pattern has been linked to what researchers describe as "metacognitive laziness", a decline in students' ability to monitor and assess their learning (Fan et al., 2024). Similarly, Deng et al. (2025) found that using ChatGPT in educational settings can reduce cognitive effort and lead to weaker argumentation and critical thinking. These findings suggest that when ChatGPT is used without proper guidance, it may encourage surface-level learning and interrupt students' cognitive development. For this reason, teacher involvement and the cultivation of critical thinking and digital literacy are essential to ensure that GenAI supports, rather than undermines, the learning process. This distinction points to a central theme of the present review: the dual role of GenAI as either a metacognitive facilitator enhancing reflective and critical learning or a superficial answer generator, which risks undermining authentic cognitive engagement.

2.4. GenAI and ChatGPT in Science Education

Within the context of science education, ChatGPT can offer immediate responses, clarifications, and support in experimental settings, helping students bridge the gap between theory and practice (Javaid et al., 2023; Rasul et al., 2023). For instance, when working with electrical circuits, learners can adjust variables and request explanations, deepening their understanding of scientific principles such as Ohm's law. Additionally, interaction with the tool may foster analytical and critical thinking, as students engage with multiple perspectives and develop arguments (Chinonso et al., 2023).

The significance of this study lies in the growing presence of GenAI tools in education, and particularly in science teaching. While platforms like ChatGPT are becoming more widespread, empirical research on their use, especially among university students and in the teaching of abstract scientific concepts remains limited. Although some studies have begun to address this gap, there is still a lack of robust evidence evaluating ChatGPT in authentic learning environments and cognitively demanding subjects such as the natural sciences (Kotsis, 2024). At the same time, it is increasingly important for educational systems to prepare both teachers and learners to use such tools through pedagogically grounded and ethically informed practices (İpek et al., 2023; Spasopoulos et al., 2025).

3. Research Questions

Following the theoretical foundations discussed above, there is a notable gap in the literature concerning the integration of GenAI tools specifically within laboratory-based or experimental contexts in Science and STEM education. While prior research has highlighted the general potential of GenAI in educational settings, few empirical studies have systematically examined its cognitive, pedagogical, and ethical implications in hands-on scientific learning environments. The research questions formulated in this study emerged after a preliminary review of the existing literature and are closely aligned with the overall aims of the research. They also respond to broader educational challenges regarding the meaningful integration of GenAI in experimental settings. They are informed by cognitive and pedagogical outcomes, students' and educators' attitudes and perceptions, as well as the comparative evaluation of these perspectives. In response to these gaps, the present study seeks to explore the following research questions.

1) What cognitive and pedagogical benefits do students gain from the use of GenAI tools in laboratory-based or experimental courses in Science or STEM education?

2) What are the attitudes and perceptions of students and educators regarding the use of GenAI in laboratory-based or experimental Science or STEM courses?

3) What ethical dilemmas or concerns arise from the use of GenAI in laboratory or experimental Science or STEM learning environments?

4. Methodology

To form a specific and comprehensive understanding of the use of GenAI in teaching science in laboratory-based courses, the methodology of a Systematic Literature Review (SLR) was employed, following systematic and predefined steps that directly address the study's research questions.

The model used for organizing and presenting the methodology was the

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework (Page et al., 2021). The process included the following steps, as visually summarized in **Figure 1** below.

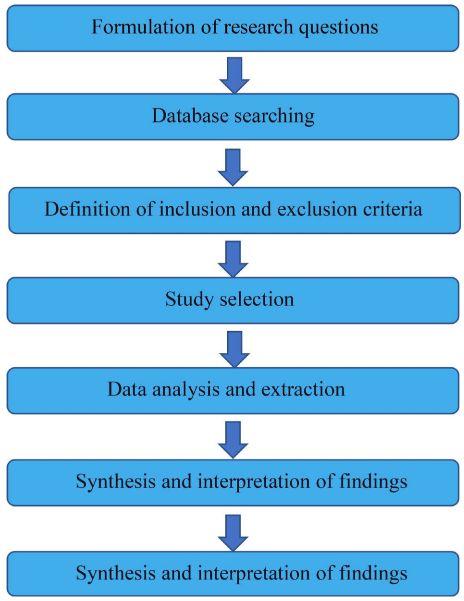

Figure 1. Stages of the systematic review process.

4.1. Search Strategy

This (SLR) focused on a carefully selected set of scholarly sources identified through academic databases as well as in the academic search engine Google Scholar. To support a robust and credible analysis, the review included only peer-reviewed journal articles and studies published in the proceedings of recognized international conferences. The databases utilized were Scopus, SpringerLink, Taylor & Francis Online, Elsevier, Wiley Online Library, ERIC, SAGE Journals, JSTOR, and Google Scholar, which was selected for its capacity to retrieve studies that are not necessarily hosted in education-specific journals, functioning as the most comprehensive academic search engine.

During the construction of search strings, it was considered that GenAI tools are referred to in the literature using various terms and synonyms. For this reason, the thesaurus of each database was used where available, and Boolean operators (AND, OR) were applied to combine the search terms. This strategy aimed to optimize the sensitivity and completeness of the review by identifying all relevant published works, as recommended in the international literature.

The terms selected for the review reflect the study's central focus areas. For the concept of GenAI, used keywords such as: GenAI, large language models, LLMs and ChatGPT. For science and STEM education, the following terms were used: science education, STEM education, physics education, chemistry education, biology education, natural sciences, scientific literacy, disciplinary teaching in science, and inquiry-based science education. Regarding laboratory and experimental learning, the search included terms such as laboratory-based learning, experimental science teaching, inquiry-based experiments, hands-on activities, wet labs, virtual labs, project-based STEM learning, lab environment, experimental practice, and exploratory learning. In terms of educational level and participants, the following terms were used: higher education, university students, undergraduate students, graduate students, teacher education, preservice teachers, pre-service science teachers, and college-level STEM courses.

All the above terms were used in multiple combinations, depending on the features and filters of each database. The search focused on studies published between 2023 and 2025 that met the predefined inclusion criteria and addressed the research questions of the present study (Table 1).

Table 1. Core concepts and search synonyms.

Core Concepts	Synonyms
Generative Artificial Intelligence (GenAi)	generative artificial intelligence, GenAI, large language models, LLMs, ChatGPT, AI-generated content, generative models
Science/STEM Education	science education, STEM education, physics education, chemistry education, biology education, natural sciences, scientific literacy, science teaching, teaching of science, STEM instruction
Laboratory/Experimental Learning	laboratory-based learning, experimental science teaching, hands-on science education, inquiry-based experiments, lab environments, wet labs, virtual labs, project-based learning, exploratory learning
Higher Education/Participants	higher education, university education, tertiary education, undergraduate students, graduate students, college-level education, preservice teachers, teacher education

After testing and reviewing the specific syntax required by each electronic database, a unified search string was constructed using Boolean operators and parentheses. The search string employed included the following core terms and their synonyms: ("generative artificial intelligence" OR "GenAI" OR "large language models" OR "LLMs" OR "ChatGPT" OR "AI-powered tools" OR "AI-generated content" OR "AI-based tutor" OR "conversational agents" OR "generative models") AND ("science education" OR "STEM education" OR "physics education" OR "chemistry education" OR "biology education" OR "natural sciences" OR "scientific literacy" OR "science teaching" OR "teaching of science" OR "STEM instruction") AND ("laboratory-based learning" OR "experimental science teaching" OR "hands-on science education" OR "inquiry-based experiments" OR "lab environments" OR "virtual labs" OR "project-based learning" OR "exploratory learning") AND ("higher education" OR "university education" OR "tertiary education" OR "undergraduate students" OR "graduate students" OR "college-level education" OR "preservice teachers" OR "teacher education").

4.2. Inclusion and Exclusion Criteria

Specific inclusion and exclusion criteria were established to evaluate and select the appropriate studies for inclusion in this SLR. These criteria enabled the selection of only those articles that were relevant to the research questions and aligned with the study's thematic and methodological framework.

Inclusion Criteria:

1) The study had to be empirical (quantitative, qualitative, or mixed methods) and conducted in an authentic learning context.

2) There had to be active use of GenAI (such as ChatGPT or other GenAI applications) within the context of teaching or learning.

3) The instruction had to take place within laboratory-based or experimental Science or STEM courses, with an emphasis on inquiry-based or project-based activities.

4) Participants had to be undergraduate or postgraduate students involved in Science Education and/or pre-service science teachers.

5) The publication had to be a peer-reviewed scientific article.

6) The publication had to be released between 2023 and 2025 to reflect the recent application of GenAI in education.

Exclusion Criteria:

1) Studies not written in English.

2) Theoretical or conceptual articles that do not include empirical data.

3) Technological evaluations of tools without pedagogical implementation or student participation in a learning context.

4) Articles that refer exclusively to non-experimental or non-laboratory environments.

5) Publications that are duplicates across databases or exist only as abstracts.

4.3. Screening, Filtering, and Final Study Selection

Due to the large number of initial results from the databases, additional filtering criteria such as publication date and logical relevance were applied, following the recommendations of Haddaway et al. (2015). For each search, up to 300 results per database were examined based on relevance. In cases where the number of results was smaller, all articles were reviewed.

The initial search yielded a total of 3400 articles. After removing duplicates and screening for publication date, title, and abstract, 3236 articles were excluded. The remaining 164 articles were assessed for their relevance to the focus of the present review.

Out of the 164 articles, 92 were excluded because they did not fully meet the inclusion criteria, such as lacking a laboratory implementation or active integration of GenAI tools. As a result, 72 articles were examined in depth concerning the research questions and selection criteria.

The process involved thematic analysis, during which each study was read repeatedly, with detailed notes taken on methodology, interventions, and findings, followed by comparative evaluation across the studies. Ultimately, an additional 52 studies were excluded due to insufficient empirical evidence or an exclusive focus on theoretical or technological issues, without student participation in laboratory or experimental settings.

The final sample of the systematic review includes 20 empirical studies that meet all the inclusion criteria and are analyzed in the following chapter. The selection process is summarized in the PRISMA flow diagram (Figure 2 & Table 2).

Table 2. Number of articles selected from databases.

Database	First stage Identification	Second stage Screening	Third stage eligibility	Fourth stage eligibility	Articles Included
ERIC	300	102	23	14	0
Elsevier	300	61	21	15	1
Scopus	300	36	17	13	4
SpringerLink	300	78	29	20	8
SAGE Journals	300	53	11	5	0
Taylor & Francis Online	300	92	26	18	1
JSTOR	300	71	8	2	0
Google Scholar	1000	309	11	5	4
Wiley	300	105	18	15	2
Total	3400	907	164	72	20

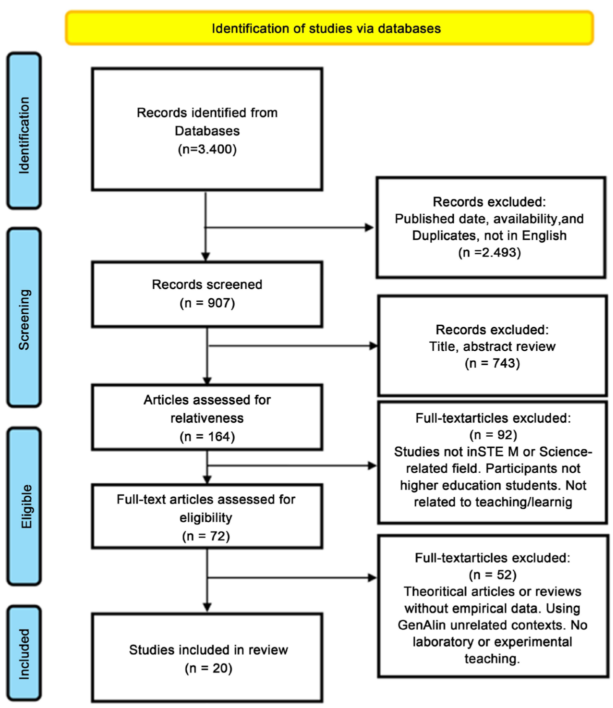

Figure 2. PRISMA flow diagram.

4.4. Data Extraction and Analysis

The selected articles were analyzed based on predefined questions and categories to extract qualitative and quantitative data related to the use of GenAI (such as ChatGPT) in experimental or laboratory-based teaching practices in higher education. The analysis focused on four principal axes: the subject area (STEM/Science Education), the level of study (undergraduate or postgraduate), the type of educational activity (laboratory, simulation, or experimental teaching), and the reported cognitive, pedagogical, affective, or ethical outcomes (Petousi & Sifaki, 2020). To enhance the validity and reliability of the process, the database source of each article was recorded, and triangulation and comparative coding procedures were applied across studies. The structure and criteria of this analysis are visually summarized in **Table 3**. In the Appendix (**Table A1**) please find the titles of the 20 articles.

Table 3. Articles included in the SLR.

No	Author(s)	Educational Level	Content Area	Educational Context
1	Kim et al. (2024)	Higher (UG)	**Electrical Engineering** and **Mechanical Engineering** (STEM)	Students used ChatGPT to revise lab reports in engineering courses. The study compared original and revised reports to assess improvements in writing quality.
2	Lin et al., (2024)	Higher (UG)	STEM Education	Students used a GPT-based chatbot as a **learning aid** during self-directed study. The chatbot provided real-time support, explanations, and feedback. The study used **multimodal learning analytics** to examine students' engagement and learning behaviors.
3	West et al. (2023)	Higher (UG)	Chemistry Education	Students used **ChatGPT** to generate full or partial **chemistry lab reports** across different levels of the curriculum. The study analyzed the AI-generated reports for accuracy, depth, and alignment with course expectations and examined students' **perceptions** of ChatGPT's usefulness and limitations.
4	Blonder et al. (2024)	Higher (UG pre-service science teachers)	Science Education (Chemistry)	Pre-service chemistry teachers used ChatGPT to design and reflect on chemistry lessons. The focus was on how ChatGPT-supported tasks revealed their **Pedagogical Content Knowledge** (PCK) and supported their teacher preparation.
5	Niloy et al. (2024)	Higher (UG)	Computer Science Education (STEM)	Students participated in an experimental course where **ChatGPT** acted as a **co-teacher** alongside the instructor. ChatGPT provided real-time explanations, feedback, and help during lessons and assignments. The study compared learning outcomes and perceptions between groups with and without ChatGPT support.
6	Ravi et al. (2025)	Higher (UG - Engineering students)	Engineering Education (STEM)	Students and educators co-designed and used **LLM-based tools (e.g., ChatGPT)** to support **project-based learning** in engineering courses. Tools helped students with ideation, feedback, and documentation during team projects.
7	Matzakos & Moundridou (2025)	Higher (UG - Engineering students)	Engineering Education (STEM)	Students used **ChatGPT** to solve computational problems and to support report writing in engineering assignments. The study explored students' usage patterns, perceptions, and the impact on academic performance.

Continued

8	Sirisathitkul et al. (2025)	Higher (UG)	Physics Education	Students engaged in **inquiry-based learning** using **ChatGPT** to explore the physics of microwave ovens. ChatGPT served as a **conversational assistant**, providing answers, explanations, and clarifications throughout the experimental process.
9	Choi (2025)	Higher (UG)	Earth Science Education	Students used **ChatGPT** alongside **Earth science simulations** to analyze natural phenomena. The AI provided explanations, supported conceptual understanding, and assisted students in interpreting simulation data during inquiry-based activities.
10	Nascimento et al. (2024)	Higher (UG)	Chemistry Education	Students used a **multi-modal AI system** combining **ChatGPT, DALL·E, and image interpretation tools** to support their understanding of chemistry concepts. The AI assisted in generating textual explanations, visual representations, and interpreting diagrams to enhance students' conceptual learning.
11	Lin et al. (2025)	Higher (UG)	STEM Education	Students engaged in **experiential STEM activities** supported by **ChatGPT** as a reflective learning aid. The AI was used to prompt students' reflection, provide feedback, and guide them in articulating and deepening their understanding during hands-on learning tasks.
12	Liang et al. (2023)	Higher (UG)	Physics Education	Students used **ChatGPT** during physics learning tasks to ask conceptual questions, solve problems, and receive real-time explanations. The study focused on students' interaction patterns with the AI, its influence on understanding, and perceived effectiveness in supporting individual learning.
13	Fanelli et al. (2024)	Higher UG	Physics Education	Pre-service physics teachers used ChatGPT to question, evaluate, and reflect on physics concepts and explanations. The study focused on developing their didactic reasoning and critical engagement with AI tools during teacher preparation.
14	Ji et al. (2025)	Higher UG	STEM Education	Students collaborated with **ChatGPT** during a STEM course focused on entrepreneurship. The AI was used to **co-create solutions**, support **critical thinking**, and enhance **learning performance**. The study measured effects on **AI awareness**, **cognitive load**, and students' **perceptions** of the human–machine learning process.
15	Elkhodr et al. (2023)	Higher UG	Information and Communication Technology (STEM)	Students participated in **reflective lab sessions** where they interacted with **ChatGPT** to complete ICT-related tasks. The study focused on students' **perceptions**, **ethical concerns**, and reflections on using GenAI as part of their lab-based learning.
16	El Fathi et al. (2025)	Higher (UG – First-year engineering students)	STEM Education	Students used **ChatGPT** to support **conceptual understanding** and correct **scientific misconceptions** during STEM learning tasks. The study also assessed students' **acceptance** and **perceptions** of GenAI as a learning tool.

Continued

17	Beltozar-Clemente & Díaz-Vega (2025)	Higher UG	Physics Education	Students in a Physics 1 course used **ChatGPT** combined with **gamification elements** (such as interactive quizzes and badges) to enhance their **academic performance** and **learning motivation**. The tool provided **guidance**, **explanations**, and **feedback** during experimental activities, and its effectiveness was evaluated through students' **performance outcomes** and **attitudinal responses**.
18	Khodadad (2025)	Higher UG	Physics Education	Students critically evaluated pre-generated ChatGPT responses to physics questions. The activity aimed to foster critical thinking by having them assess scientific accuracy, clarity, and potential misconceptions in AI-generated content
19	Ahmed et al. (2024)	Higher UG	Computer Science (STEM)	Students in an introductory programming course used **ChatGPT** as a **virtual teaching assistant**. The tool provided explanations, debugging help, and code suggestions. The study evaluated ChatGPT's **effectiveness**, **viability**, and **students' perceptions** regarding its use in learning programming concepts.
20	Polverini & Gregorcic (2024)	Higher UG	Physics Education	Students conducted **experimental physics tasks** and used **ChatGPT** to generate code for **data acquisition and analysis** with Arduino-based setups. The study explored how ChatGPT supported **problem-solving**, **conceptual understanding**, and students' **interaction with measurement technology** in a lab environment.

A systematic keyword-based search was conducted to identify and collect empirical studies that met the predefined inclusion criteria. To ensure the reliability and consistency of the extracted data, a triangulation process was applied, which included initial coding, cross-verification across sources, and final document analysis (Nowell et al., 2017). From each article, the following key information was systematically extracted to support the comparative synthesis of findings:

1) Subject area
2) Educational context (e.g., laboratory, simulation, project-based environment)
3) Methodology
4) Theoretical framework or learning models
5) Learning outcomes (cognitive, affective, or skills-based)
6) Pedagogical outcomes (e.g., self-regulation, inquiry support, feedback)
7) Participants' attitudes and perceptions toward GenAI tools
8) Type and mode of interaction with AI tools
9) Assessment tools used (e.g., performance tests, questionnaires, interviews)

5. Findings

Before analyzing the main research questions, the three most frequently occurring keywords in the selected studies were identified in order to highlight the dominant themes that were developed and analyzed. The three most prevalent keywords were: GenAI, science education, and pedagogical benefits. These initial findings suggest that the primary orientation of the literature is to highlight the pedagogical value of integrating GenAI tools into the context of science teaching, mainly in terms of supporting conceptual understanding, inquiry-based learning, and student engagement in experimental or laboratory activities.

5.1. Methodology and Evaluation Tools

The studies included in this (SLR) employ a diverse range of methodological approaches, with a predominance of quantitative and mixed-method designs, alongside a notable presence of qualitative research (6 out of 20 studies). Quantitative studies employ standardized instruments, such as comprehension tests, attitude questionnaires, and performance measurements, providing objective and comparable data on the effectiveness of using GenAI in science education. The mixed-method studies, by combining quantitative and qualitative tools, enable triangulation of findings, thus enhancing the validity of conclusions and shedding light on both cognitive and affective dimensions of the student experience (Creswell et al., 2011). Equally noteworthy is the contribution of qualitative research, which provides in-depth insights into students' interactions with GenAI tools through content analysis, interviews, and observation of tools such as ChatGPT in real learning environments. The methodological diversity of the studies included in this review enables a multifaceted analysis of AI integration in higher education, focusing not only on student performance, but also on perceptions, emotions, and the pedagogical dimensions of their learning experiences. (**Table 4**)

Table 4. Methodological approach and data collection tools of the studies.

No.	Author(s)	Methodology	Data Collection Tools
1	Kim et al. (2024)	Mixed methods	Student lab reports, focus group interview, rubric-based textual analysis
2	Lin et al. (2024)	Mixed methods	Learning management system logs, self-report questionnaires, semi-structured interviews
3	West et al. (2023)	Mixed methods	lab reports, Student surveys, Interviews
4	Blonder et al. (2024)	Mixed methods	Likert-scale questionnaire, Open-ended questions, Descriptive statistics, Content analysis
5	Niloy et al. (2024)	Mixed methods	Pre-test, post-test, Likert-scale survey, and interviews
6	Ravi et al. (2025)	Qualitative	Interviews
7	Matzakos & Moundridou (2025)	Qualitative	Observations, interviews, student feedback

Continued

8	Sirisathitkul et al. (2025)	Qualitative	Video recordings, student reflections, classroom dialogue transcripts
9	Choi (2025)	Mixed methods	Pre- and post-tests, classroom observations, student interview
10	Nascimento et al. (2024)	Qualitative	Prompt-based student tasks and ChatGPT responses
11	Lin et al. (2025)	Mixed methods	Self-reflection worksheets, observation logs, interviews
12	Liang et al. (2023)	Qualitative	Open-ended questionnaire and observation of student discussions
13	Fanelli et al. (2024)	Quantitative	Survey questionnaires, performance scores
15	Ji et al. (2025)	Mixed methods	Pretest and posttest questionnaires, statistical tests, systematic review of reflective learning reports, and analysis of chat records.
16	Elkhodr et al. (2023)	Mixed methods	Observation notes, Student reflective exercises, Rubric scores assessing student work on tutorial tasks
17	El Fathi et al. (2025)	Mixed methods	Pre-test, post-test, and Likert-scale questionnaire
18	Beltozar-Clemente & Díaz-Vega (2025)	Mixed methods	Pre-test, post-test, and questionnaire on student motivation
18	Khodadad (2025)	Qualitative	Analysis of student-designed prompts and ChatGPT-generated responses
19	Ahmed et al. (2024)	Mixed methods	Pre-test, post-test, and student perception survey
20	Polverini & Gregorcic (2024)	Mixed methods	Questionnaire, Open-ended questions, Performance-based assessment

The analysis of the data collection tools used in the 20 studies included in this review reveals diversity in both methodological approaches and collection techniques. In the mixed-method studies, which constitute the majority, the most common tools were pre- and post-tests, Likert-type questionnaires, semi-structured interviews, as well as laboratory reports and content analysis of student dialogues or written work. The quantitative studies primarily relied on self-report questionnaires, performance tests, and statistical analysis of scores, aiming to measure changes in knowledge, attitudes, and motivation levels.

In the qualitative studies, data were collected through interviews, observations, video analysis, reflective worksheets, and content generated with ChatGPT, including reflective responses, dialogues, and student-designed prompts. Some studies also utilized data from digital platforms, such as digital observation logs or Learning Management System (LMS) records. The variety of data collection tools enhances the validity of the findings, allowing the investigation of GenAI use at cognitive, pedagogical, and emotional levels.

These elements are schematically represented in **Figure 3** below, which visualizes the structure and sources of qualitative data used across the reviewed studies.

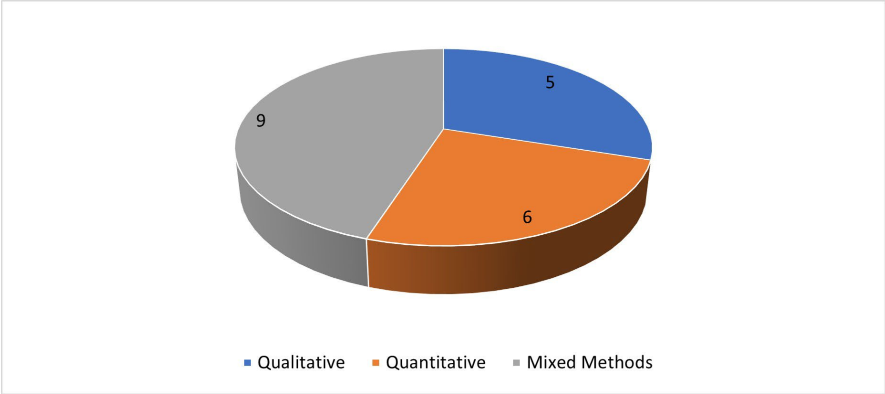

Figure 3. Data collection.

Regarding data collection techniques, most studies collected data from multiple sources. The most frequently used tool was questionnaires (14 studies), followed by tests and assessments (9), interviews (8), students' reflective or open-ended responses (5), and observation methods such as field notes or video recordings (3). Additionally, in some cases, data from digital platforms were utilized, such as gamification records or interaction logs (2).

These data are summarized in Figure 4 below, which provides a visual overview of the types and frequency of data collection methods employed across the selected studies.

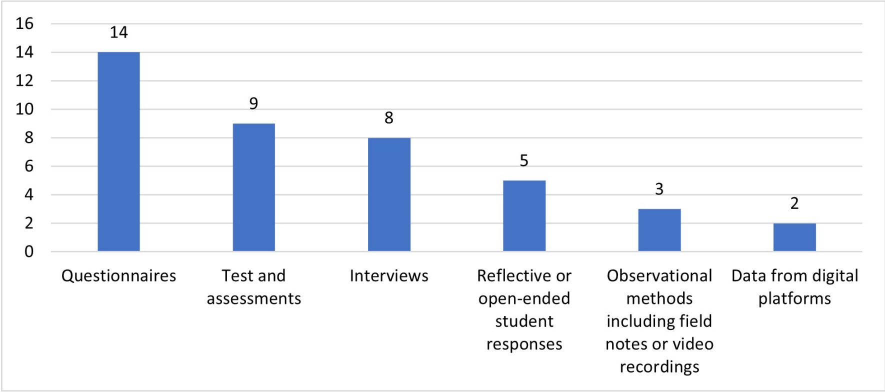

Figure 4. Assessment tools.

Based on the data collected from the 20 empirical studies included in this (SLR), all investigations were conducted in university settings, involving undergraduate or postgraduate students. Regarding the subject area, nine studies focused on Physics (9), four on Chemistry (4), two on Biology (2), two on Computer Science (2), while three studies (3) involved interdisciplinary or broader STEM applications, incorporating various concepts from the natural sciences and technology.

The educational context of the analyzed studies is characterized by an applied, experimental, and inquiry-based orientation, as all were implemented in higher education environments. The interventions primarily consisted of activities centered on learning by doing and problem-solving. In 12 studies, students participated in laboratory exercises (either physical or virtual) in 9 studies, they engaged in solving scientific problems or analyzing experimental data and in 7 cases, project-based instructional scenarios or simulations were implemented. As visually depicted in **Figure 5** all studies reported active student engagement with GenAI tools primarily ChatGPT, which were used as integrally embedded supports in the learning process. Specifically, in 10 studies, GenAI was used for writing laboratory reports in 9 studies, for solving scientific problems or analyzing experimental data in 7 studies, for generating scientific explanations and in 9 studies, for guiding the execution of experiments. It is worth noting that several studies incorporated more than one of these uses, highlighting the flexibility of GenAI tools and their capacity to support the learning process holistically within experimental settings.

Based on the analysis of the 20 empirical studies included in the SLR, it was found that all studies (20/20) utilized ChatGPT. In five of these studies, ChatGPT was used in combination with other GenAI tools. Specifically, DALL·E was used in two studies for image generation, Whisper in one study for audio transcription, GitHub Copilot in one study for programming environments, Bing Chat (Microsoft Copilot) in one study, and Google Gemini (Bard) in one study.

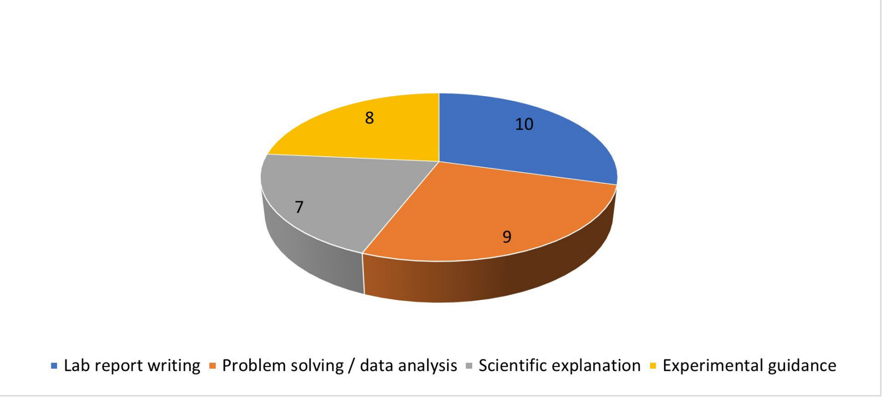

Figure 5. Forms of student interaction with GenAI.

5.2. Theoretical Framing of GenAI Integration in the Experimental Process

The design of the interventions identified in the studies included in this SLR was not explicitly based on specific pedagogical models. However, their content and structure reflect theoretical orientations drawn from contemporary approaches in educational science. The active engagement of students, hypothesis formulation, experimental verification, and reflective interpretation of results suggest the adoption of constructivist principles, where knowledge is constructed through experience and personal reconstruction (McGuire et al., 2024). The use of tools such as ChatGPT within experimental and inquiry-based activities aligns with the stages of Kolb's experiential learning cycle (Kolb, 2014), which includes concrete experience, reflective observation, abstract conceptualization, and active experimentation all of which are evident in the descriptions of the instructional scenarios.

Moreover, several interventions incorporate digital interactions that resonate with the principles of Vygotsky's sociocultural theory, as ChatGPT is often used as a "cognitive scaffold" operating within the zone of proximal development, enhancing the learning process through dialogue and personalized guidance. From another perspective, Connectivism (Siemens, 2005) appears particularly relevant to the way students connect information, questions, and digital tools to construct knowledge in learning networks, with ChatGPT functioning as a hub in this flow of information. Finally, Cognitive Load Theory (Sweller et al., 2011) provides interpretive tools for understanding the role of GenAI as a regulatory mechanism of cognitive load, especially in complex tasks such as scientific report writing or data analysis. Recent research by Yan et al. (2025) confirms that the use of GenAI reduces extraneous load, thereby supporting a focus on learning without bypassing essential knowledge acquisition.

Overall, the findings suggest that even in the absence of explicit theoretical framing most applications align with multifactorial theoretical principles that support self-regulated, inquiry-based, and digitally mediated learning. It should be noted that the authors of the reviewed studies did not explicitly state these theoretical alignments, but instead represent the reviewers' interpretation based on the pedagogical features and design elements identified across the interventions.

Following the analytical review of the empirical studies included in the (SLR), recurring findings were identified and categorized into five core areas: cognitive outcomes, pedagogical benefits, students' attitudes and perceptions, teachers' perspectives, and ethical dilemmas arising from the integration of GenAI in educational practice. These categories appear systematically across the reviewed articles and form the foundational structure for interpreting the results.

To facilitate comparative analysis and the extraction of conclusions, the following summary table of findings was developed. The results of the review were coded into thematic categories using unique abbreviations. Each finding corresponds to a distinct code indicating its category and sequence number. Additionally, for enhanced readability and interpretability, findings associated with challenges or difficulties are highlighted in red within the table.

The table below summarizes the main findings of the 20 empirical studies included in this (SLR), presenting the types of outcomes identified based on a predefined coding scheme (**Table 5**). The presentation is organized into five main categories: cognitive outcomes, pedagogical implications, students' attitudes and perceptions, teachers' attitudes and perceptions, and ethical concerns. Each cell of the table includes the corresponding finding codes observed in each study, offering a comparative overview of the thematic focus and coverage of the selected literature. This structure facilitates the identification of common patterns and recurring trends across the studies.

Table 6 below summarizes the main findings of the 20 empirical studies included in this (SLR), presenting the types of outcomes identified based on a

Table 5. Coded findings from the reviewed articles.

Cognitive outcomes		
C1: Improved conceptual understanding	C2: Enhanced problem-solving skills	C3: Increased accuracy in scientific reasoning
C4: Correct application of scientific principles	C5: Improved lab report quality and structure	C6: Better performance in quizzes/tests
C7: Longer retention of knowledge	C8: Higher-order thinking	
Pedagogical outcomes		
P1: Supported inquiry-based learning	P2: Scaffolded experimentation	P3: Teaching assistant/mentor role
P4: Facilitated flipped/blended learning	P5: Enhanced collaboration	P6: Helped structure investigations
P7: Enabled differentiated instruction	P8: Smooth instructional integration	
Students' attitudes & perceptions		
S1: Found helpful for learning	S2: Increased confidence in experiments	S3: Higher satisfaction with learning
S4: Concerns about overreliance on GenAI	S5: Felt more autonomous	S6: Useful for revision/feedback
S7: Perceived as reliable support tool		
Teachers' attitudes & perceptions		
T1: Valuable classroom assistant	T2: Concern for academic integrity	T3: Recognized personalization potential
T4: Emphasized pedagogical framing	T5: Concern over student dependence on AI	T6: Viewed as complement, not replacement
Ethical concerns/dilemmas		
E1: Academic dishonesty/plagiarism	E2: Unclear authorship	E3: Overreliance on GenAI content
E4: Lack of transparency about GenAI limitations	E5: Fairness in assessment	E6: Need for ethical guidelines

predefined coding scheme (Table 6). The presentation is organized into five main categories: cognitive outcomes, pedagogical implications, students' attitudes and perceptions, teachers' attitudes and perceptions, and ethical concerns. Each cell of the table includes the corresponding finding codes observed in each study, offering a comparative overview of the thematic focus and coverage of the selected literature. This structure facilitates the identification of common patterns and recurring trends across the studies.

5.3. Cognitive Findings

The findings from the 20 selected studies consistently demonstrate the enhancement of cognitive outcomes through the integration of GenAI tools, such as ChatGPT, in laboratory and experimental STEM settings. The most frequently

Table 6. Categorization and coding of research findings.

Articles	Cognitive Findings	Pedagogical Findings	Students' attitudes and perceptions	Teachers' attitudes and perceptions	Ethical Considera-tions
Kim et al. (2024)	**C1, C4**	**P2, P3**	S1, S2, S6	T3, T4	E1, E2, E3
Lin et al. (2024)	C1, C2, C6, C8	P2, P3, P6, P8	S1, S3, S6	Not mentioned	E2, E3
West er al. (2024)	**C3, C4, C5**	**P3, P6, P8**	**S1, S2, S6**	Not mentioned	**E1, E2, E3**
Blonder et al. (2024)	C1, C2, C8	P3, P6, P7,	S1, S6, S4	T3, T4	E1, E2
Niloy et al. (2024)	C1, C2, C3, C4, C6, C7, C8.	P1, P2, P3, P4, P5, P6, P8	S1, S2, S3, S5, S6, S7, S4	T1, T3, T4, T6, T5	E1, E3, E4, E6
Ravi et al. (2025)	C1, C2, C4, C8	P1, P2, P3, P4, P6, P7, P8	S1, S2, S3, S5, S6, S7, S4	T1, T3, T4, T6	E3, E4, E6
Matzakos & Moundridou (2025)	C1, C2, C3, C4, C5, C6, C8	P1, P2, P3, P5, P6, P8	S1, S3, S6, S7, S4	T1, T4, T6	E1, E2, E3, E6
Sirisathitkul et al. (2025)	C1, C2, C3, C4, C8	P1, P2, P3, P6, P8	S1, S3, S6, S7, S4	T4, T6	E3, E6
Choi (2025)	C1, C2, C3, C4, C5, C8	P1, P2, P3, P6, P8	S1, S3, S6, S7, S4	T1, T3, T4, T6	E3, E6
Nascimento et al. (2024)	C1, C3, C4, C8	Prompt engineering supported differentiation and complexity, P3, P8	S4	T4, T2, T5	E1, E3, E4, E6
Lin et al. (2025)	C1, C2, C8	P1, P2, P3, P8	S1, S3, S4, S6, S7, S4	T1, T3, T4	E2, E3, E6
Liang et al. (2023)	C1, C2, C3, C4	P1, P2, P3, P6, P8	S1, S3, S6, S7, S4	T1, T4, T6	E2, E3
Fanelli et al. (2024)	C1, C2, C4, C8	P1, P2, P3, P8	S1, S2, S3, S6, S7, S4	T1, T4	E2, E3, E6
Ji et al. (2025)	C1, C2, C4, C5, C6, C7, C8	P1, P2, P3, P6, P8	S1, S2, S3, S6, S7, S4	T1, T3, T4, T6	E2, E3, E4, E6
Elkhodr et al. (2023)	C1, C2, C4, C8	P1, P3, P6, P7	S1, S2, S3, S6, S7, S4	Not mentioned	E2, E3, E6
El Fathi et al. (2025)	C1, C2, C3, C4, C6, C8	P1, P2, P3, P6, P8	S1, S2, S3, S6, S7	T1, T3, T4, T6	E2, E3, E4, E6
Beltozar-Clemente & Díaz-Vega (2025)	C1, C2, C3, C6, C8	P1, P3, P6, P8	S1, S3, S6, S7	Not mentioned	E3, E6
Khodadad (2025)	C1, C2, C3, C6	P1, P2, P6	S1, S3, S6	Not mentioned	E3, E6
Ahmed et al. (2024)	C1, C2, C4, C8	P1, P2, P3, P6, P8	S1, S3, S6, S7, S4	T1, T3, T4, T5	E2, E3, E6
Polverini & Gregorcic (2024)	C1, C4, C6	P1, P2, P5, P6	S1, S2, S6	T4, T6	Not mentioned

reported outcome is improved conceptual understanding (C1, 20 occurrences), which, as noted in the study by Holmes et al. (2023), reflects the ability of generative AI tools to support conceptual transitions from theoretical to practical levels.

The correct application of scientific principles (C4, 18 occurrences) is also frequently mentioned, a finding corroborated by Nguyen and Hayward's (Nguyen and Hayward, 2025) study, which documents increased accuracy in students' use of scientific terminology. The enhancement of problem-solving skills (C2, 16 occurrences) is equally prevalent, as confirmed by Zhou et al. (2024), who report that the ease of using GenAI tools is positively associated with students' problem-solving abilities, as it facilitates the application of self-regulation strategies and allows focus on complex cognitive processes.

The development of higher-order thinking (C8, 14 occurrences) is supported by Wang and Fan (2025), who found that the use of ChatGPT enhanced students' analytical and conceptual processing skills during the design of instructional interventions. Increased accuracy in the use of scientific concepts (C3, 10 occurrences) is further supported by the findings of Nguyen and Hayward (2025). Additionally, improvements in test performance (C6, 10 occurrences) are documented in the study by Lin et al. (2025), which reports a positive impact of GenAI use on science assessment scores. The quality of laboratory reports (C5, 4 occurrences) is also highlighted by Zhou et al. (2024), who found that students produced more detailed and well-supported reports when using AI tools.

Finally, long-term knowledge retention (C7, 2 occurrences) is supported by the findings of Sharma et al. (2025), which indicate that integrating LLM tools into assessment is associated with enhanced stability in recalling scientific concepts. These results are visually summarized in **Figure 6** which represents the frequency and distribution of the reported cognitive outcomes across the selected studies.

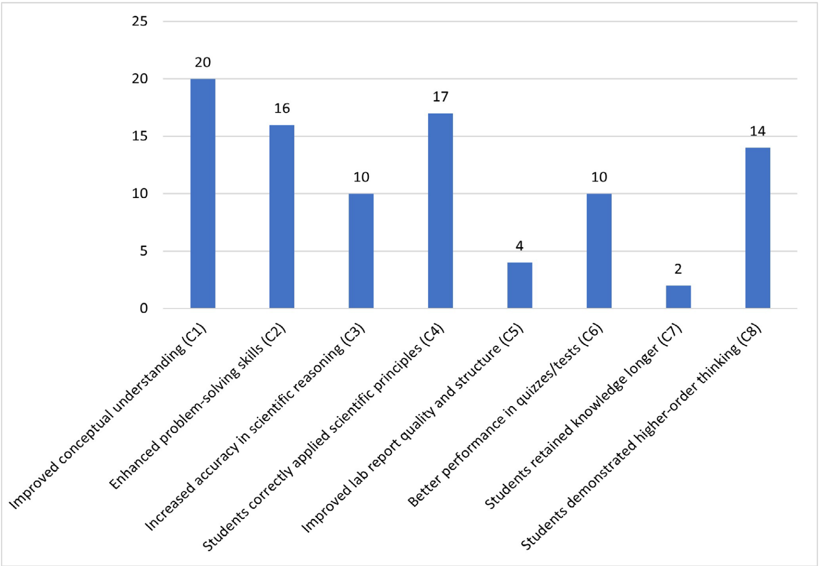

Figure 6. Cognitive findings.

5.4. Pedagogical Findings

The analysis of the pedagogical findings revealed a broad range of opportunities for utilizing GenAI in experimental and laboratory-based STEM instruction. A key finding was the role of GenAI as a teaching assistant or mentoring presence, recorded in 19 instances (P3), confirming the expectation that such technology can function as a personalized learning companion. In a similar context, van den Berg (2024) describes the use of ChatGPT by educators to support translation, assessment, and instructional activities, enhancing everyday teaching practices.

Guidance during the experimental process (16 instances, P2) also emerged as a significant pedagogical function of GenAI. Gonçalves Costa et al. (2024) demonstrated that instructional scenarios designed with GPT-4 offered greater organizational coherence and pedagogical scaffolding compared to traditional approaches. Likewise, support for inquiry-based learning (16 instances, P1) was prominently recorded. Nguyen and Hayward (2025) documented the contribution of GenAI in fostering reflective interaction and interpreting scientific phenomena through assessment-driven strategies.

An important finding was the facilitation of students' research organization (18 instances, P6), which aligns with the observations of Kaplan-Rakowski et al. (2023), who emphasize that the use of GenAI by educators is associated with increased organizational efficiency and support for student learning processes. The seamless integration of GenAI into the instructional flow (13 instances, P8) was also highlighted in several studies. Nikolic et al. (2025) note that, under appropriate assessment design and learning objectives, the technology can be integrated creatively without disrupting instructional coherence.

In fewer cases, the facilitation of differentiated instruction was observed in four instances (P7). However, the study by Fuller and Barnes (2024) highlights that GenAI can support personalized learning by offering adaptability to students' characteristics and needs. Support for hybrid or flipped classroom models (4 instances, P4) is also documented in the study by Chiu (2024), who proposes a model for integrating GenAI into self-regulated learning environments aligned with the phases of forethought, performance, and reflection.

Finally, the promotion of collaborative learning (4 instances, P5) appears to be facilitated by GenAI, as demonstrated by Chan and Hu (2023), who emphasize the role of the technology as a mediating tool for group interaction and cognitive exchange. Taken together, these findings underscore the importance of pedagogical focus and guidance for the effective integration of GenAI into teaching practice.

Figure 7 below offers a visual summary of the pedagogical dimensions identified across the selected studies.

5.5. Students' Attitudes and Perceptions

Students' attitudes toward the use of GenAI tools were predominantly positive, with multidimensional indicators of engagement, acceptance, and satisfaction. The most frequently reported finding was the recognition of the tool's usefulness for feedback and review (S6, 19 occurrences), as confirmed by the study of Sinha et al. (2024), in which students used the BoilerTAI platform to receive targeted support in forum discussions and reflective activities, thereby deepening their understanding of the content.

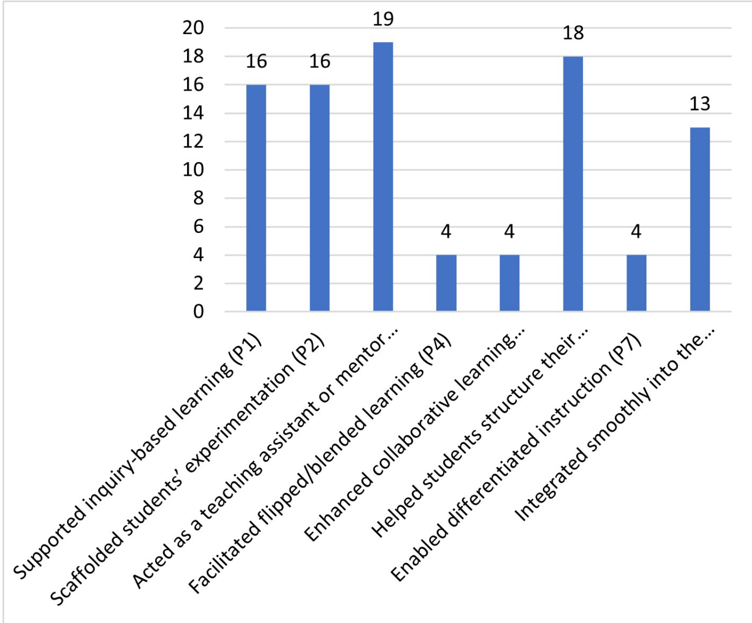

Figure 7. Pedagogical findings.

The contribution of GenAI to the learning process (S1, 19 occurrences) is also documented in Tang (2024), who highlights that GenAI meaningfully supports conceptual processing, particularly in linguistically mediated environments. Satisfaction with the overall experience (S3, 15 occurrences) is reinforced by the findings of Pack et al. (2024), who report high levels of student acceptance and satisfaction with the use of GenAI tools in writing tasks.

The perception of GenAI as a reliable support mechanism (S7, 15 occurrences) is confirmed by Blahopoulou & Ortiz Bonnin (2025), where the majority of students expressed trust in the answers provided, while remaining aware of the need for critical evaluation. Increased confidence during experimental activities (S2, eight occurrences) is also reported in the same study, with students stating that the use of GenAI offered them a sense of support and safety.

The heightened sense of autonomy in learning (S5, 2 occurrences) is supported by Qian (2025), in which students utilized GenAI to independently explore concepts and understand complex terminology, thereby fostering their ability to proceed without immediate instructor intervention.

Finally, concerns about overreliance on artificial intelligence (S4, 11 occurrences) are highlighted by both Zhai et al. (2024) and Sinha et al. (2024), with students expressing fear that constant use might hinder their independence and development of critical thinking skills.

The diverse student perspectives discussed above are collectively depicted in **Figure 8**, providing a visual summary of their prevalence.

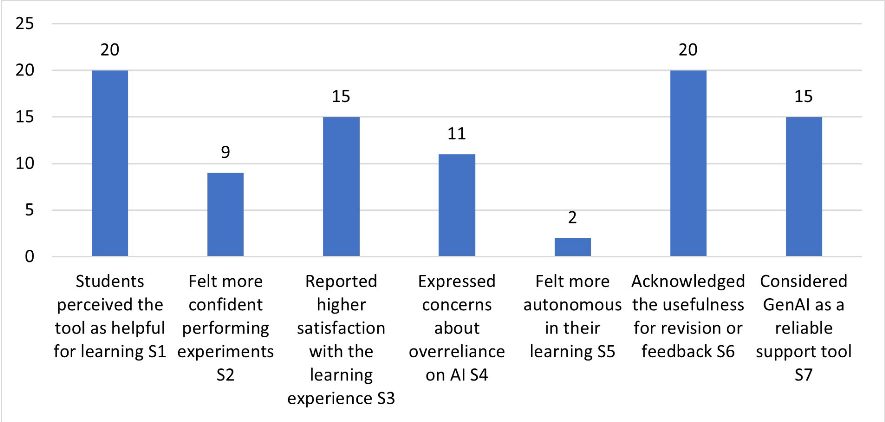

Figure 8. Findings on students' attitudes and perceptions.

5.6. Teachers' Attitudes and Perceptions

The teachers who participated in the studies reviewed expressed a range of views regarding the integration of GenAI tools in experimental and laboratory-based teaching. The most frequently mentioned issue was the need for pedagogical framing of the tool's use (T4, 16 occurrences), as spontaneous or unstructured use is considered insufficient to achieve learning objectives. This concern is also confirmed in the study by Bower et al. (2024), in which educators emphasized the need for clear guidance on how to incorporate ChatGPT within pedagogical frameworks.

At the same time, educators recognized GenAI's potential to support personalized learning (T3, 10 occurrences), a finding substantiated by the study of van den Berg et al. (2024), which highlights the flexibility of GenAI to adapt to the pace and learning style of different students. An equal number of studies documented positive acceptance of the tool as a valuable assistant within the teaching process (T1), consistent with the findings of Kaplan-Rakowski et al. (2023), who emphasize GenAI's contribution to instructional preparation, assessment, and guidance.

In ten cases, it was noted that GenAI can function as a complementary, not competitive, presence alongside the instructor (T6), suggesting the establishment of a supportive technological role. This is also emphasized by Nikolic et al. (2025), who warn against the risk of pedagogically inappropriate automation if roles are not clearly defined.

However, some reservations were also noted: in three studies, concerns were raised about students' potential overreliance on GenAI (T5), and in one study, teachers expressed concerns about academic integrity (T2), particularly regarding the transparency of the tool's contribution to final student work. These concerns are echoed in a study by Bower et al. (2024), which notes that the uncritical use of ChatGPT may blur the line between authentic and AI-assisted work.

Overall, teachers' attitudes appear to combine a high level of acceptance with a conscious awareness of the boundaries and conditions that should govern the pedagogical use of GenAI.

A summary of these perceptions and concerns is illustrated in **Figure 9**, which categorizes the teachers' responses based on frequency and thematic relevance across the selected studies.

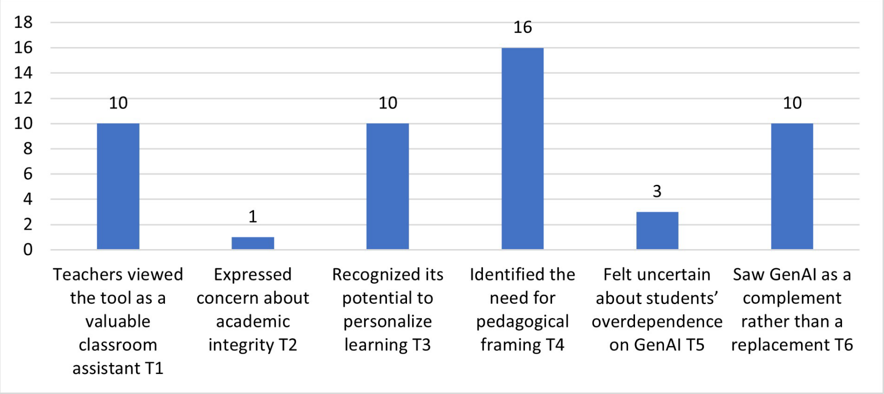

Figure 9. Findings on teachers' attitudes and perceptions.

5.7. Ethical Issues and Dilemmas

Despite the pedagogical and learning benefits of integrating GenAI tools in higher education, several studies highlighted critical ethical issues that require careful pedagogical management. The most frequently reported concern was students' overreliance on GenAI-generated content (E3, 19 occurrences), raising risks regarding diminished autonomy and critical thinking (Shahzad & Xu, 2024). The issue of verifying the extent of student participation in the production of submitted work (E2, 14 occurrences) raises serious questions of transparency and assessment fairness (Tang, 2024).

Additionally, 15 studies emphasized the need for the formulation of official ethical guidelines (E6) concerning the use of GenAI tools in educational contexts (Mittal et al., 2024). The lack of adequate student awareness about the limitations of GenAI tools (E4, six occurrences), along with concerns about potential violations of academic integrity (E1, five occurrences), further reinforces the need for clear pedagogical and ethical practices (Dekerlegand et al., 2025; Ogunleye et al., 2024).

Finally, issues related to fairness in assessment (E5), although identified in only one study, underscore the importance of ensuring equitable treatment of students when GenAI technologies are used - an issue also linked to academic integrity principles (Ogunleye et al., 2024).

These findings underscore the need for a clearly defined regulatory and value-based framework to accompany the integration of GenAI in education, thereby safeguarding the fundamental principles of academic integrity and meaningful learning.

Figure 10 provides a visual synthesis of the key ethical concerns identified in the reviewed studies, illustrating the frequency and distribution of each issue across the selected literature.

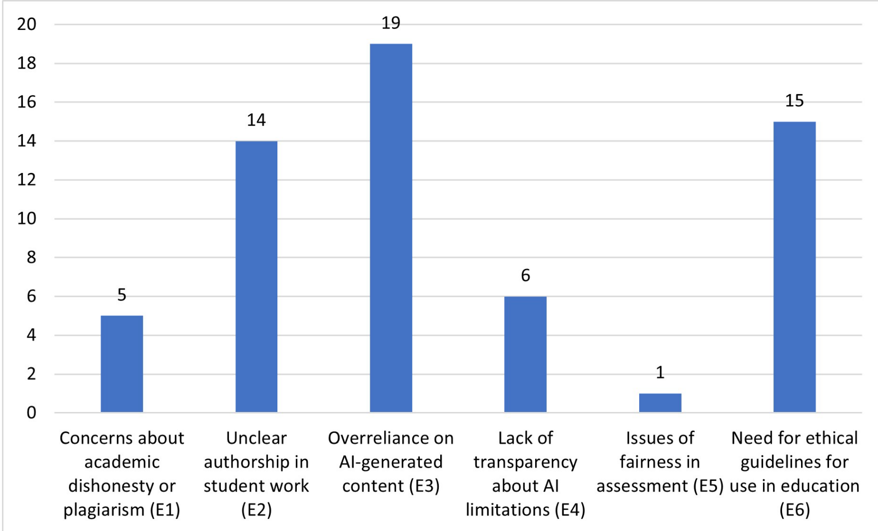

Figure 10. Findings on ethical issues and dilemmas.

6. Discussion

In this section, the findings and conclusions of the review are discussed about each of the predefined research questions. To ensure a clear and systematic interpretation, the discussion is structured according to the research questions and is explicitly linked to the thematic categories that emerged during the analysis: Cognitive Outcomes (C), Pedagogical Outcomes (P), Students' Attitudes and Perceptions (S), Teachers' Attitudes and Perceptions (T), and Ethical Issues and Dilemmas (E). This structure facilitates targeted reflection and allows for a more coherent synthesis of the results.

6.1. Answering Research Question 1: Cognitive and Pedagogical Outcomes (C, P)

Addressing the first research question regarding the cognitive benefits of using GenAI tools in experimental STEM and science education environments, recent studies demonstrate that tools such as ChatGPT can significantly enhance conceptual understanding. For instance, in the experimental study by Niloy et al. (2024), the integration of ChatGPT as a "co-teacher" allowed students to autonomously explore part of the curriculum with AI support, resulting in deeper learning, improved conceptual grasp, higher performance in assessments, and increased learner satisfaction. Similarly, in introductory computer science courses, Ahmed et al. (2024) found that ChatGPT effectively explained fundamental programming concepts, helping first-year students understand the core principles of the field. Additionally, in a university-level physics course that combined ChatGPT with gamification strategies, Beltozar-Clemente and Díaz-Vega (2024) reported that the experimental group outperformed the control group in terms of academic achievement. These students also developed a more robust understanding of physical phenomena, as reflected in their ability to solve more complex problems successfully. In engineering education, Kim et al. (2024) noted that the use of

ChatGPT in writing laboratory reports improved report quality and helped students gain a deeper understanding of academic writing conventions and structure. Collectively, these findings suggest that GenAI integration can support conceptual clarity and the internalization of difficult subject matter in science and STEM education, functioning as a supplementary tool to promote cognitive development (Uğraş et al., 2025).

Regarding the pedagogical benefits, which also fall under the first research question, the use of GenAI is associated with promoting learner autonomy and reflective thinking, contributing to self-regulated learning. In a laboratory course for engineering students, ChatGPT functioned as a support tool that encouraged students to take a more active role in the learning process. As a result, students demonstrated increased self-direction and metacognitive awareness, identifying knowledge gaps and adjusting their learning strategies accordingly. Using GenAI as a complementary rather than sole source of knowledge can promote self-regulated learning without replacing students' critical thinking, as noted by Elkhodr et al. (2023). These findings suggest that the responsible integration of GenAI can empower learners by fostering autonomy and enhancing their metacognitive functions.

6.2. Answering Research Question 2: Students' and Teachers' Perceptions (S, T)

Concerning the second research question, which examines students' perceptions of GenAI integration, the positive attitudes displayed by students towards the use of GenAI in laboratory and experimental coursework lend strong support to the encouraging findings previously discussed. Students consistently acknowledged its role in enhancing their learning experiences and fostering greater active participation. Specifically, in science and computer science contexts, students perceived ChatGPT as a supportive tool that promoted conceptual understanding, strengthened autonomy, and provided immediate feedback, making experimental tasks more accessible and comprehensible (Polverini & Gregorcic, 2024; Fanelli et al., 2024). In engineering lab environments, Ravi et al. (2025) and Niloy et al. (2024) reported that ChatGPT was beneficial for writing lab reports, providing guidance and enhancing the clarity of scientific language. Moreover, students appreciated AI as a "second thought" partner that fostered reflection and supported their reasoning processes, as highlighted by Blonder et al. (2024) and (Uğraş et al., 2025).

This second research question also encompasses educators' similarly positive views. Teachers saw GenAI as a valuable complementary tool that could assist in providing individualized student support. Particularly in the context of their professional development, many educators described their interaction with ChatGPT as an opportunity to reflect on their Pedagogical Content Knowledge (PCK), noting that it helped them better understand their students' learning challenges (Blonder et al., 2024; Zhou et al., 2024; Uğraş et al., 2025). Across the reviewed studies, it was evident that the overall atmosphere of acceptance was powerful when GenAI was framed as a tool that enhances rather than replaces the educator's role.

6.3. Answering Research Question 3: Ethical Dilemmas and Limitations (E)

About the third research question, which concerns the ethical dilemmas associated with the use of GenAI in higher STEM education, the literature highlights that, despite its clear benefits, the integration of tools such as ChatGPT also raises several significant ethical concerns. One of the most prominent issues is the safeguarding of academic integrity, as GenAI tools can facilitate instances of plagiarism or unauthorized assistance. For example, Elkhodr et al. (2023) documented instances where students used AI-generated responses to dishonestly complete their assignments.

Another primary concern relates to the potential for misinformation. Since ChatGPT can generate inaccurate or unverifiable content without citing sources, students may struggle to assess the validity of the information they receive (Kim et al., 2024). Furthermore, GenAI models may contain algorithmic biases, producing outputs that systematically favor certain perspectives or social groups, which could inadvertently reinforce inequities in educational settings.

Lastly, many researchers caution against students' overreliance on such tools. When used excessively or without critical engagement, GenAI may reduce learners' autonomy and hinder the development of independent thinking skills. As noted by Choi (2025), there is a growing risk that students may become passive recipients of information, relying too heavily on AI-generated answers rather than engaging in active learning processes.

6.4. Emerging Tensions and Contradictions Across Findings

A closer synthesis of the findings derived from the research questions reveals several characteristic tensions between findings. For example, the cognitive enhancement reported across numerous studies, particularly regarding higher-order thinking, problem-solving, and scientific understanding (C8, C2, C1), appears to be challenged by student and educator concerns about overreliance on GenAI tools (S4, E3), which may limit the development of autonomous thinking and meaningful cognitive engagement. Similarly, although educators acknowledge the potential of GenAI to support instruction and personalized learning (T1, T3), they simultaneously emphasize the need for pedagogical framing (T4) and a clear role definition between teacher and technological tool (T6), to prevent its use in pedagogically inadequate ways. Finally, while students frequently report increased confidence and a sense of support when using such tools (S2, S7), this positive experience is not always accompanied by sufficient awareness of the technology's limitations (E4), raising further concerns about the need to strengthen digital literacy and metacognitive awareness.

7. Conclusions

7.1. Addressing Research Question 1: Cognitive and Pedagogical Benefits

This review highlighted that the integration of GenAI tools, such as ChatGPT, into laboratory-based science education can meaningfully enhance students' conceptual understanding and overall academic performance. Among the key cognitive benefits identified were deeper engagement with core scientific concepts, improved application of scientific principles, and higher scores on evaluation tasks.

From a pedagogical perspective, GenAI tools were found to promote student autonomy, activate metacognitive processes, and encourage learners to take an active role in their learning journey. Students were not merely seeking answers, but using the tool to explore areas of uncertainty, reflect on their thinking, and adapt their learning strategies accordingly.

Overall, the findings suggest that when GenAI tools are embedded within pedagogically grounded frameworks, they can complement traditional instruction. These tools have the potential to facilitate comprehension, interpretation, and application of scientific knowledge, supporting both the cognitive and pedagogical dimensions of learning.

7.2. Addressing Research Question 2: Students' and Teachers' Perceptions of GenAI Use

The analysis of empirical findings revealed a generally positive attitude among both students and educators toward the use of GenAI tools, especially in laboratory and experimental contexts. Students particularly in disciplines such as science, computer science, and engineering viewed ChatGPT as a supportive resource for conceptual understanding, problem-solving, and promoting autonomy during experimental tasks. Its ability to provide immediate feedback was seen as a key factor in enhancing learning and increasing engagement.

At the same time, educators did not perceive GenAI as a threat, but rather as a tool that can enhance their teaching practice. Some even described their engagement with ChatGPT as a valuable opportunity for reflecting on their pedagogical content knowledge (PCK), helping them better identify students' learning gaps and instructional needs. Acceptance was notably higher when GenAI was framed as a collaborative assistant rather than a potential replacement for the teacher's role.

Nevertheless, several concerns were voiced by students. Some reported feelings of overreliance on the tool or difficulty distinguishing their understanding from the information provided by GenAI. Others expressed doubts about the accuracy of specific AI-generated responses, as well as concerns about reduced interaction with their instructors. These reservations do not negate the overall positive perception but instead underscore the importance of pedagogically structured GenAI use, with clearly defined roles and objectives.

The analysis of students' and educators' perspectives confirms the added value of GenAI in educational settings, provided it is used with critical awareness and appropriate guidance, in ways that support learner autonomy and active engagement without undermining the importance of human interaction in the educational process (Spasopoulos et al., 2025).

7.3. Addressing Research Question 3: Ethical Dilemmas in GenAI-Supported Experimental Learning

The SLR identified a range of ethical challenges emerging from the integration of GenAI tools in higher STEM education. A central issue was the protection of academic integrity, as the automatic generation of content increases the risk of plagiarism and creates ambiguity about the authenticity of student work.

Concerns were also raised regarding misinformation. GenAI can sometimes produce responses that contain inaccuracies or lack grounding in reliable sources. This highlights the need for critical evaluation of AI-generated content, as well as pedagogical scaffolding to help students navigate and assess the information they encounter. Additionally, some GenAI models may exhibit algorithmic bias, which could lead to unequal treatment or the reinforcement of stereotypes.

Finally, the risk of overdependence on GenAI tools remains a significant concern, especially in experimental learning environments where active student participation is essential. Excessive reliance may diminish critical thinking, creativity, and learners' autonomy. In such contexts, unchecked use of GenAI could suppress curiosity and cognitive engagement, which are vital to the inquiry-driven nature of laboratory learning.

Addressing these ethical concerns does not imply restricting the use of GenAI, but rather calls for the implementation of clear educational frameworks, institutional guidelines, and the promotion of digital responsibility. Thoughtful instructional design and targeted student training can play a crucial role in fostering ethical awareness and minimizing the potential negative consequences of GenAI integration in higher education.

8. Limitations-Future Research Directions

Although this SLR, offers a multi-layered and well-substantiated account of the use of GenAI tools in the education of future science teachers, it is subject to certain limitations that must be considered when interpreting the findings.

First, the exclusive selection of peer-reviewed, English-language scientific articles may introduce both publication bias and linguistic bias. Studies with inconclusive or less impactful results are often left unpublished or underrepresented, which may affect the balance of the overall conclusions. Additionally, focusing solely on English-language sources may have led to the omission of significant contributions from other scientific communities.

Second, substantial heterogeneity was observed among the included studies in terms of disciplinary focus, research design, sample size, and educational context.

This variability complicates the process of drawing generalizable conclusions and limits the external validity of the synthesis.

Third, there was an uneven distribution of subject areas, with most studies focusing on physics and computer science. In contrast, fields such as biology, environmental sciences, and educational sciences were underrepresented or absent. Furthermore, core concepts such as "cognitive enhancement," "metacognition," and "pedagogical integration" were interpreted differently across studies, making it challenging to formulate thematically focused comparisons or recommendations.

Finally, many of the studies featured limitations related to small sample sizes and short intervention durations, further constraining the ability to derive long-term or statistically robust conclusions (Kanaki & Kalogiannakis, 2023).

This review identified significant benefits from the use of GenAI in science education; however, several areas warrant further investigation. First, there is a need for longitudinal studies to assess the long-term impact of tools such as ChatGPT on student learning. Most existing studies focus on short-term interventions, leaving unclear whether the observed positive effects persist over time.

Furthermore, it is recommended that the pedagogical integration of GenAI be strengthened through the development of appropriate instructional strategies and frameworks. The effectiveness of these tools largely depends on how they are integrated into teaching, which underscores the need for effective implementation models in problem-solving scenarios, laboratory activities, or project-based learning (PBL) environments.

Additionally, research should be expanded to include underrepresented subject areas. Specifically, future studies should focus on underrepresented fields such as Biology, Chemistry, and Environmental Science in order to broaden the empirical evidence base and explore how GenAI tools can support diverse scientific domains. Similarly, the perspectives of educators remain insufficiently explored, despite their crucial role in the successful integration of AI tools into teaching practice.

Ultimately, further research is necessary to investigate the role of GenAI in fostering the development of scientific thinking and laboratory skills. Future investigations could examine how ChatGPT supports hypothesis generation, data analysis, and student reflection without replacing students' reasoning. Suggested directions include the use of guided instructional scenarios or simulated experiments to strengthen problem-solving competencies.

Funding

The authors received no financial support for this article's research and/or authorship.

Ethics Declaration

All procedures followed were under the ethical standards of the Ethical Committee of the University of Thessaly, Greece.

Declaration of Interest

The author declares no competing interest.

Data Availability

Data generated or analyzed during this study are available from the authors on request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Ahmed, Z., Shanto, S. S., & Jony, A. I. (2024). Potentiality of Generative AI Tools in Higher Education: Evaluating ChatGPT's Viability as a Teaching Assistant for Introductory Programming Courses. *STEM Education, 4,* 165-182. https://doi.org/10.3934/steme.2024011

Ai, Y., Baveja, M., Girdhar, A., O'Dell, M., & Deorio, A. (2024). A Custom Generative AI Chatbot as a Course Resource. In *2024 ASEE Annual Conference & Exposition Proceedings* (pp. 1-14). American Society for Engineering Education. https://doi.org/10.18260/1-2--46433

Ausat, A. M. A., Massang, B., Efendi, M., Nofirman, & Riady, Y. (2023). Can ChatGPT Replace the Role of the Teacher in the Classroom: A Fundamental Analysis. *Journal on Education, 5,* 16100-16106. http://jonedu.org/index.php/joe/article/view/2745

Beltozar-Clemente, S., & Díaz-Vega, E. (2025). Physics XP: Integration of ChatGPT and Gamification to Improve Academic Performance and Motivation in Physics 1 Course. *International Journal of Engineering Pedagogy, 14,* 82-92. https://doi.org/10.3991/ijep.v14i6.47127

Blahopoulou, J., & Ortiz-Bonnin, S. (2025). Student Perceptions of ChatGPT: Benefits, Costs, and Attitudinal Differences between Users and Non-Users toward AI Integration in Higher Education. *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13575-9

Blonder, R., Feldman-Maggor, Y., & Rap, S. (2024). Are They Ready to Teach? Generative AI as a Means to Uncover Pre-Service Science Teachers' PCK and Enhance Their Preparation Program. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-024-10180-2

Bower, M., Torrington, J., Lai, J. W. M., Petocz, P., & Alfano, M. (2024). How Should We Change Teaching and Assessment in Response to Increasingly Powerful Generative Artificial Intelligence? Outcomes of the ChatGPT Teacher Survey. *Education and Information Technologies, 29,* 15403-15439. https://doi.org/10.1007/s10639-023-12405-0

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early Experiments with GPT-4*. https://doi.org/10.48550/arXiv.2303.12712

Candrasari, R., Makulua, J., Noviasmy, Y., Makulua, K., & Siminto, S. (2024). GPT Chat: Useful or Not in Supporting Learning in Higher Education. *International Journal of Language and Ubiquitous Learning, 2,* 113-125. https://doi.org/10.70177/ijlul.v2i2.963

Chan, C. K. Y., & Hu, W. (2023). Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education. *International Journal of Educational Technology* *in Higher Education, 20,* Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chinonso, O. E., Theresa, A. M., & Aduke, T. C. (2023). ChatGPT for Teaching, Learning and Research: Prospects and Challenges. *Global Academic Journal of Humanities and Social Sciences, 5,* 33-40. https://doi.org/10.36348/gajhss.2023.v05i02.001

Chiu, T. K. F. (2024). A Classification Tool to Foster Self-Regulated Learning with Generative Artificial Intelligence by Applying Self-Determination Theory: A Case of ChatGPT. *Educational Technology Research and Development, 72,* 2401-2416. https://doi.org/10.1007/s11423-024-10366-w

Choi, Y. (2025). Earth Science Simulations with Generative Artificial Intelligence. *Journal of Science Education and Technology, 34,* 215-232. https://doi.org/10.53761/nf1yqr46

Collie, R. J., & Martin, A. J. (2025). Teachers' Early Uptake of genAI in Teaching and Learning: Important Questions and Answers. *Technology, Mind, and Behavior, 28,* Article No. 93. https://doi.org/10.1007/s11218-025-10052-6

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). Sage Publications.

Dekerlegand, R., Bell, A., Clancy, M. J., Pletcher, E. R., & Pollen, T. (2025). Generative Artificial Intelligence in Education: Insights from Rehabilitation Sciences Students. *Education Sciences, 15,* Article 380. https://doi.org/10.3390/educsci15030380

Delgado, A., Wardlow, L., O'Malley, K., & McKnight, K. (2015). Educational Technology: A Review of the Integration, Resources, and Effectiveness of Technology in K-12 Classrooms. *Journal of Information Technology Education: Research, 14,* 397-416. https://doi.org/10.28945/2298

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT Enhance Student Learning? A Systematic Review and Meta-Analysis of Experimental Studies. *Computers & Education, 227,* Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Ding, L., Kim, S., & Allday, R. A. (2024). Development of an AI Literacy Assessment for Non-Technical Individuals: What Do Teachers Know? *Contemporary Educational Technology, 16,* ep512. https://doi.org/10.30935/cedtech/14619

El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahmi, E. M. (2025). Integrating Generative AI into STEM Education: Enhancing Conceptual Understanding, Addressing Misconceptions, and Assessing Student Acceptance. *Disciplinary and Interdisciplinary Science Education Research, 7,* Article No. 6. https://doi.org/10.1186/s43031-025-00125-z

Elkhodr, M., Gide, E., Wu, R., & Darwish, O. (2023). ICT Students' Perceptions towards ChatGPT: An Experimental Reflective Lab Analysis. *STEM Education, 3,* 70-88. https://doi.org/10.3934/steme.2023006

European Union (2024). Regulation (EU) *2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA Relevance).* https://data.europa.eu/eli/reg/2024/1689/oj

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y. et al. (2024). Beware of Metacognitive Laziness: Effects of Generative Artificial Intelligence on Learning Motivation, Processes, and Performance. *British Journal of Educational Technology, 56,* 489-530. https://doi.org/10.1111/bjet.13544

Fanelli, C., Giroux, J., Moran, P., Nayak, H., Suresh, K., & Walter, E. (2024). Using Generative AI Tools to Support Data Analysis in Laboratory Work. *Journal of Instrumenta-*

tion, 19, C07011. https://doi.org/10.1088/1748-0221/19/07/C07011

Fuller, M., & Barnes, N. (2024). The Impact of ChatGPT on Teaching and Learning in Higher Education: Exploring the Dual Perspectives of Participants Who Were Students and Teachers. *New Directions for Higher Education, 2024,* 31-46. https://doi.org/10.1002/he.20507

Gonçalves Costa, G., J. D. Nascimento Júnior, W., Mombelli, M. N., & Girotto Júnior, G. (2024). Revisiting a Teaching Sequence on the Topic of Electrolysis: A Comparative Study with the Use of Artificial Intelligence. *Journal of Chemical Education, 101,* 3255-3263. https://doi.org/10.1021/acs.jchemed.4c00247

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLOS ONE, 10,* e0138237. https://doi.org/10.1371/journal.pone.0138237

Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' Technological Pedagogical Content Knowledge and Learning Activity Types. *Journal of Research on Technology in Education, 41,* 393-416. https://doi.org/10.1080/15391523.2009.10782536

Holmes, W., Bialik, M., & Fadel, C. (2023). *Artificial Intelligence in Education*. Globethics Publications. https://doi.org/10.58863/20.500.12424/4276068

Huang, J., Saleh, S., & Liu, Y. (2021). A Review on Artificial Intelligence in Education. *Academic Journal of Interdisciplinary Studies, 10,* 206-217. https://doi.org/10.36941/ajis-2021-0077

İpek, Z. H., Gözüm, A. İ. C., Papadakis, S., & Kalogiannakis, M. (2023). Educational Applications of the ChatGPT AI System: A Systematic Review Research. *Educational Process International Journal, 12,* 26-55. https://doi.org/10.22521/edupij.2023.123.2

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the Opportunities through ChatGPT Tool Towards Ameliorating the Education System. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3,* Article 100115. https://doi.org/10.1016/j.tbench.2023.100115

Ji, Y., Zhan, Z., Li, T., Zou, X., & Lyu, S. (2025). Human–Machine Cocreation: The Effects of ChatGPT on Students' Learning Performance, AI Awareness, Critical Thinking, and Cognitive Load in a STEM Course toward Entrepreneurship. *IEEE Transactions on Learning Technologies, 18,* 402-415. https://doi.org/10.1109/tlt.2025.3554584

Kanaki, K., & Kalogiannakis, M. (2023). Sample Design Challenges: An Educational Research Paradigm. *International Journal of Technology Enhanced Learning, 15,* 266-285. https://doi.org/10.1504/ijtel.2023.10055808

Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., & Papin, K. (2023). Generative AI and Teachers' Perspectives on Its Implementation in Education. *Journal of Interactive Learning Research, 34,* 313-338. https://doi.org/10.70725/815246mfssgp

Kent, C., & du Boulay, B. (2022). *AI for Learning*. CRC Press. https://doi.org/10.1201/9781003194545

Khodadad, D. (2025). ChatGPT in Engineering Education: A Breakthrough or a Challenge? *Physics Education, 60,* Article 045006. https://doi.org/10.1088/1361-6552/add073

Kim, D., Majdara, A., & Olson, W. (2024). A Pilot Study Inquiring into the Impact of ChatGPT on Lab Report Writing in Introductory Engineering Labs. *International Journal of Technology in Education, 7,* 259-289. https://doi.org/10.46328/ijte.691

Kolb, D. A. (2014). *Experiential Learning: Experience as the Source of Learning and Development* (2nd ed.). Pearson Education.

Kotsis, K. T. (2024). ChatGPT in Teaching Physics Hands-On Experiments in Primary School. *European Journal of Education Studies, 11,* Article 5549.

https://doi.org/10.46827/ejes.v11i10.5549

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,* 1643-1660. https://doi.org/10.1109/tlt.2024.3401457

Liang, Y., Zou, D., Xie, H., & Wang, F. L. (2023). Exploring the Potential of Using ChatGPT in Physics Education. *Smart Learning Environments, 10,* Article No. 52. https://doi.org/10.1186/s40561-023-00273-7

Lin, C. J., Lee, H. Y., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Enhancing Reflective Thinking in STEM Education through Experiential Learning: The Role of Generative AI as a Learning Aid. *Education and Information Technologies, 30,* 6315-6337. https://doi.org/10.1007/s10639-024-13072-5

Lin, C. J., Wang, W. S., Lee, H. Y., Li, P. H., Huang, Y. M., & Wu, T. T. (2025). Advancing Self-Directed Learning in STEM Education: Integrating GPT-Based Learning Aid with Multimodal Learning Analytics. *Journal of Research on Technology in Education,* 1-19.

Matzakos, N., & Moundridou, M. (2025). Exploring Large Language Models Integration in Higher Education: A Case Study in a Mathematics Laboratory for Civil Engineering Students. *Computer Applications in Engineering Education, 33,* e70049. https://doi.org/10.1002/cae.70049

McGuire, A., Qureshi, W., & Saad, M. (2024). A Constructivist Model for Leveraging Genai Tools for Individualized, Peer-Simulated Feedback on Student Writing. *International Journal of Technology in Education, 7,* 326-352. https://doi.org/10.46328/ijte.639

Mittal, U., Sai, S., Chamola, V., & Sangwan, D. (2024). A Comprehensive Review on Generative AI for Education. *IEEE Access, 12,* 142733-142759. https://doi.org/10.1109/access.2024.3468368

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the Implementation of ChatGPT in Education: A Systematic Review. *Computers, 12,* Article 153. https://doi.org/10.3390/computers12080153

Nascimento, M. D. R., Barreto, J. L. M., Nascimento, A. G. B., Silva, P. D. C. L., & de Araújo, I. C. M. (2024). Enhancing AI Responses in Chemistry: Integrating Text Generation, Image Creation, and Image Interpretation through Different Levels of Prompts. *Journal of Chemical Education, 101,* 3767-3779. https://doi.org/10.1021/acs.jchemed.4c00230

Nguyen, A., Kremantzis, M., Essien, A., Petrounias, I., & Hosseini, S. (2024). Editorial: Enhancing Student Engagement through Artificial Intelligence (AI): Understanding the Basics, Opportunities, and Challenges. *Journal of University Teaching and Learning Practice, 21,* 1-13. https://doi.org/10.53761/caraaq92

Nguyen, H., & Hayward, J. (2025). Applying Generative Artificial Intelligence to Critiquing Science Assessments. *Journal of Science Education and Technology, 34,* 199-214. https://doi.org/10.1007/s10956-024-10177-x

Nikolic, S., Suesse, T. F., Grundy, S., Haque, R., Lyden, S., Lal, S. et al. (2025). Assessment Integrity and Validity in the Teaching Laboratory: Adapting to Genai by Developing an Understanding of the Verifiable Learning Objectives behind Laboratory Assessment Selection. *European Journal of Engineering Education, 50,* 673-701. https://doi.org/10.1080/03043797.2025.2456944

Niloy, A. C., Ghosh, M., Alghowinem, S., Shirmohammadi, S., & Mirsamadi, S. (2024). Can Generative AI Be an Effective Co-Teacher? An Experiment. *Education and Information Technologies, 29,* 5561–5585. https://doi.org/10.1016/j.caeai.2025.100418

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods, 16,* 1-13. https://doi.org/10.1177/1609406917733847

OECD (2023). *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*. OECD Publishing. https://doi.org/10.1787/c74f03de-en

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher Education Assessment Practice in the Era of Generative AI Tools. *Journal of Applied Learning & Teaching, 7,* 1-11. https://doi.org/10.37074/jalt.2024.7.1.28

OpenAI (2024). *GPT-4 System Card. OpenAI.* https://cdn.openai.com/papers/gpt-4-system-card.pdf

Öztop, F. (2023). A Meta-Analysis of the Effectiveness of Digital Technology-Assisted STEM Education. *Journal of Science Learning, 6,* 136-142. https://doi.org/10.17509/jsl.v6i2.52316

Pack, A., Barrett, A., & Escalante, J. (2024). Large Language Models and Automated Essay Scoring of English Language Learner Writing: Insights into Validity and Reliability. *Computers and Education: Artificial Intelligence, 6,* Article 100234. https://doi.org/10.1016/j.caeai.2024.100234

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D. et al. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *British Medical Journal, 372,* n71. https://doi.org/10.1136/bmj.n71

Petousi, V., & Sifaki, E. (2020). Contextualizing Harm in the Framework of Research Misconduct. Findings from a Discourse Analysis of Scientific Publications. *International Journal of Sustainable Development, 23,* 149-174. https://doi.org/10.1504/IJSD.2020.10037655

Polverini, G., & Gregorcic, B. (2024). How Understanding Large Language Models Can Inform the Use of ChatGPT in Physics Education. *European Journal of Physics, 45,* Article 025701. https://doi.org/10.1088/1361-6404/ad1420

Qian, Y. (2025). Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field. *TechTrends*. https://doi.org/10.1007/s11528-025-01100-1

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences, 13,* Article 5783. https://doi.org/10.3390/app13095783

Rasul, T., Nair, S., Kalendra, D., Robin, M. et al. (2023). The Role of ChatGPT in Higher Education: Benefits, Challenges, and Future Research Directions. *Journal of Applied Learning and Teaching, 6,* 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Ravi, S., Srivastava, V., Mishra, S., & Dey, A. (2025). Co-Designing Large Language Model Tools for Project-Based Learning with K12 Educators. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-25). https://doi.org/10.1145/3706598.3713971

Shahzad, M. F., Xu, S., & Zahid, H. (2024). Exploring the Impact of Generative AI-Based Technologies on Learning Performance through Self-Efficacy, Fairness & Ethics, Creativity, and Trust in Higher Education. *Education and Information Technologies, 30,* 3691-3716. https://doi.org/10.1007/s10639-024-12949-9

Sharma, S., Mittal, P., Kumar, M., & Bhardwaj, V. (2025). The Role of Large Language Models in Personalized Learning: A Systematic Review of Educational Impact. *Discover Sustainability, 6,* Article 243. https://doi.org/10.1007/s43621-025-01094-z

Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology and Distance Learning, 2,* 3-10. http://www.itdl.org/Journal/Jan_05/article01.htm

Sinha, A., Goyal, S., Sy, Z., Kuperus, R., Dickey, E., & Bejarano, A. (2024). BoilerTAI: A Platform for Enhancing Instruction Using Generative AI in Educational Forums. In *2024 IEEE Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE. https://doi.org/10.1109/fie61694.2024.10893137

Sirisathitkul, C., Sirisathitkul, Y., Wattanasit, K., & Nisoa, M. (2025). Converging Objects-to-Think-with: Exploring Microwave Ovens through Inquiry-Based Learning with ChatGPT. *Research in Science & Technological Education,* 1-17. https://doi.org/10.1080/02635143.2025.2496990

Spasopoulos, T., Sotiropoulos, D., & Kalogiannakis, M. (2025). Generative AI in Pre-Service Science Teacher Education: A Systematic Review. *Advances in Mobile Learning Educational Research, 5,* 1501-1523. https://doi.org/10.25082/amler.2025.02.007

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Tang, K. S. (2024). Informing Research on Generative Artificial Intelligence from a Language and Literacy Perspective: A Meta-Synthesis of Studies in Science Education. *Science Education, 108,* 1329-1355. https://doi.org/10.1002/sce.21875

Timotheou, S., Miliou, O., Dimitriadis, Y., Sobrino, S. V., Giannoutsou, N., Cachia, R. et al. (2023). Impacts of Digital Technologies on Education and Factors Influencing Schools' Digital Capacity and Transformation: A Literature Review. *Education and Information Technologies, 28,* 6695-6726. https://doi.org/10.1007/s10639-022-11431-8

Ugras, H., Ugras, M., Papadakis, S., & Kalogiannakis, M. (2024). Innovative Early Childhood STEM Education with ChatGPT: Teacher Perspectives. *Technology, Knowledge and Learning, 30,* 809-831. https://doi.org/10.1007/s10758-024-09804-8

Uğraş, M., Çakır, Z., Zacharis, G., & Kalogiannakis, M. (2025). ChatGPT in Early Childhood Science Education: Can It Offer Innovative Effective Solutions to Overcome Challenges? *Computers, 14,* 368. https://doi.org/10.3390/computers14090368

United Nations Educational, Scientific and Cultural Organization (2023). *Guidance for Generative AI in Education and Research*. UNESCO. https://doi.org/10.54675/EWZM9535

Valeri, F., Nilsson, P., & Cederqvist, A. (2025). Exploring Students' Experience of ChatGPT in STEM Education. *Computers and Education: Artificial Intelligence, 8,* Article 100360. https://doi.org/10.1016/j.caeai.2024.100360

van den Berg, G. (2024). Generative AI and Educators: Partnering in Using Open Digital Content for Transforming Education. *Open Praxis, 16,* 130-141. https://doi.org/10.55982/openpraxis.16.2.640

Wang, J., & Fan, W. (2025). The Effect of ChatGPT on Students' Learning Performance, Learning Perception, and Higher-Order Thinking: Insights from a Meta-Analysis. *Humanities and Social Sciences Communications, 12,* Article No. 621. https://doi.org/10.1057/s41599-025-04787-y

West, J. K., Franz, J. L., Hein, S. M., Leverentz-Culp, H. R., Mauser, J. F., Ruff, E. F. et al. (2023). An Analysis of AI-Generated Laboratory Reports across the Chemistry Curriculum and Student Perceptions of ChatGPT. *Journal of Chemical Education, 100,* 4351-4359. https://doi.org/10.1021/acs.jchemed.3c00581

Williamson, B., & Eynon, R. (2020). Historical Threads, Missing Links, and Future Directions in AI in Education. *Learning, Media and Technology, 45,* 223-235. https://doi.org/10.1080/17439884.2020.1798995

Wu, J., Wang, J., Lei, S., Wu, F., & Gao, X. (2025). The Impact of Metacognitive Scaffolding on Deep Learning in a Genai-Supported Learning Environment. *Interactive Learning*

Environments, 1-18. https://doi.org/10.1080/10494820.2025.2479162

Yan, L., Greiff, S., Lodge, J. M., & Gašević, D. (2025). Distinguishing Performance Gains from Learning When Using Generative AI. *Nature Reviews Psychology, 4,* 435-436. https://doi.org/10.1038/s44159-025-00467-5

Yoon, H., Hwang, J., Lee, K., Roh, K. H., & Kwon, O. N. (2024). Students' Use of Generative Artificial Intelligence for Proving Mathematical Statements. *ZDM—Mathematics Education, 56,* 1531-1551. https://doi.org/10.1007/s11858-024-01629-0

Zhai, C., Wibowo, S., & Li, L. D. (2024). The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review. *Smart Learning Environments, 11,* Article No. 28. https://doi.org/10.1186/s40561-024-00316-7

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G. et al. (2024). A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT. *International Journal of Machine Learning and Cybernetics*. https://doi.org/10.1007/s13042-024-02443-6

Zhou, L., & Li, J. (2023). The Impact of ChatGPT on Learning Motivation: A Study Based on Self-Determination Theory. *Education Science and Management, 1,* 19-29. https://doi.org/10.56578/esm010103

Appendix

Table A1. Core articles from the Systematic Literature Review (SLR).

No.	Author(s)	Title
1	Kim et al. 2024	A Pilot Study Inquiring into the Impact of ChatGPT on Lab Report Writing in Introductory Engineering Labs
2	Lin et al. 2025	Advancing self-directed learning in STEM education: integrating GPT-based learning aid with multimodal learning analytics
3	West et al. 2023	An Analysis of AI-Generated Laboratory Reports across the Chemistry Curriculum and Student Perceptions of ChatGPT
4	Blonder et al. 2024	Are They Ready to Teach? Generative AI as a Means to Uncover Pre-Service Science Teachers' PCK and Enhance Their Preparation Program
5	Niloy et al. 2024	Can generative AI Be an effective Co-Teacher? An experiment
6	Ravi et al. 2025	Co-designing Large Language Model Tools for Project-Based Learning with K12 Educators
7	Matzakos & Moundridou 2025	Exploring Large Language Models Integration in Higher Education: A Case Study in a Mathematics Laboratory for Civil Engineering Students
8	Sirisathitkul et al. 2025	Converging *objects-to-think-with*: exploring microwave ovens through inquiry-based learning with ChatGPT
9	Choi 2025	Earth Science Simulations with Generative Artificial Intelligence (GenAI)
10	Nascimento et al. 2024	Enhancing AI Responses in Chemistry: Integrating Text Generation, Image Creation, and Image Interpretation through Different Levels of Prompts
11	Lin et al. 2024	Enhancing reflective thinking in STEM education through experiential learning: The role of generative AI as a learning aid
12	Liang et al. 2023	Exploring the potential of using ChatGPT in physics education
13	Fanelli et al. 2024	Physics event classification using Large Language Models
14	Ji et al. 2025	Human–Machine Cocreation: The Effects of ChatGPT on Students' Learning Performance, AI Awareness, Critical Thinking, and Cognitive Load in a STEM Course Toward Entrepreneurship
15	Elkhodr et al. 2023	ICT students' perceptions towards ChatGPT: An experimental reflective lab analysis
16	El Fathi et al. 2025	Integrating generative AI into STEM education: enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance
17	Beltozar-Clemente & Díaz-Vega 2025	Integration of ChatGPT and Gamification to Improve Academic Performance and Motivation in Physics 1 Course
18	Khodadad 2025	ChatGPT in engineering education: a breakthrough or a challenge?
19	Ahmed et al. 2024	Potentiality of generative AI tools in higher education: Evaluating ChatGPT's viability as a teaching assistant for introductory programming courses
20	Polverini & Gregorcic 2024	How understanding large language models can inform the use of ChatGPT in physics education