Volume 4 Number 1 (2026)
January – June 2026
Page: 643-657

EDUJAVARE: International Journal of Educational Research
ISSN: 3031-2884
https://edujavare.com/index.php/edujavare
DOI: 10.70610/edujavare.v4i01.1394

Systematic Literature Review: The Effect of Generative Artificial Intelligence on Students' Critical Thinking Abilities in Science Learning

Dina Fajrin[1]
[1] Universitas Negeri Yogyakarta, Indonesia
* Correspondence e-mail; dinafajrin789@gmail.com

Article history Submitted: 2026/04/14; Revised: 2026/05/18; Accepted: 2026/06/12

Abstract

The proliferation of Generative Artificial Intelligence (GenAI) tools in education has prompted urgent questions about their impact on higher-order cognitive skills. In science education, where critical thinking and scientific reasoning are foundational, understanding how GenAI reshapes these competencies is essential. This study employs a Systematic Literature Review (SLR) following PRISMA 2020 guidelines to synthesize empirical evidence on the effects of GenAI on students' critical thinking skills in science learning contexts. A comprehensive search across Scopus, ERIC, and Google Scholar databases spanning 2020-2026 identified 25 eligible studies. Findings reveal a conditional dual impact: GenAI enhances critical thinking when integrated within structured inquiry and problem-based frameworks, but poses risks of cognitive over-reliance and surface-level processing when used without pedagogical scaffolding. Critical moderating factors include AI literacy, instructional design quality, teacher facilitation, and task structure. The novelty of this review lies in its exclusive focus on the intersection of GenAI, critical thinking, and science education. Implications for curriculum designers, science teachers, and policymakers in developing-country contexts are discussed

Keywords ChatGPT, Critical Thinking, Generative AI, HOTS, Science Education.

INTRODUCTION

The development of generative artificial intelligence (GenAI), particularly large language models (LLMs) such as ChatGPT, Google Gemini, and Microsoft Copilot, has fundamentally changed the global educational landscape (Jia et al., 2025; Zhu et al., 2026; Yusuf et al., 2024). Since ChatGPT's public launch in late 2022, adoption of this technology among students and educators has increased exponentially, transforming how students access information, build understanding, and complete academic assignments (Hon, 2025; An et al., 2026; Ren et al., 2026). In the context of science education, GenAI offers significant potential, from aiding the understanding of abstract concepts and simulating virtual experiments to providing instant feedback on student responses (Williams, 2025; Nawafleh & Al-Abbas, 2025; Fadillah et al., 2024).

However, behind this potential, a critical question remains unanswered: does the use of GenAI in science learning strengthen or weaken students' critical thinking skills? Critical thinking is a fundamental foundation of science education, encompassing argument analysis, evidence evaluation, logical reasoning, and data-driven decision-making (Facione, 1990; Paul & Elder, 2006; Anderson & Krathwohl, 2001). This set of competencies aligns with Higher-Order Thinking Skills (HOTS) in the revised Bloom's taxonomy and is closely related to scientific reasoning in science learning, namely the ability to formulate hypotheses, interpret empirical data, and draw valid conclusions (Fadillah et al., 2024; Atmojo et al., 2025; Yulanda et al., 2026).

A major concern emerging in recent literature is the risk of cognitive over-reliance on GenAI (Uyar, 2025; Zhu et al., 2026; Implom, 2025). When students receive instant answers from AI systems, they tend to bypass the deep thinking processes that are central to science learning. Implom (2025) asserts that AI without proper scaffolding has the potential to lead to surface learning and cognitive stagnation. On the other hand, several studies have shown that appropriately used AI enhances inference, creativity, and reflective thinking (Yulanda et al., 2026; Guo & Lee, 2023; Nawafleh & Al-Abbas, 2025). Majeed et al. (2025) suggest that AI has the potential to both enhance and diminish critical thinking depending on the context in which it is integrated, while Simoni et al. (2025) state that further research on the relationship between AI and critical thinking is still urgently needed

Comprehensive studies through Systematic Literature Reviews (SLR) that specifically investigate the impact of GenAI on critical thinking in science education are rare (Yusuf et al., 2024; Zhu et al., 2026; Gunsaldi et al., 2025). Existing studies tend to focus on higher education in general (Hon, 2025; Rahyuni et al., 2025) or on specific disciplines such as mathematics (Holman et al., 2025; Su et al., 2026) without simultaneously combining three key variables: GenAI, critical thinking, and science learning. This study aims to fill this gap by synthesizing empirical evidence from global scientific literature and providing implications that can be applied in the Indonesian educational context.

Studies on AI in education have grown rapidly, especially since 2023 (Rahyuni et al., 2025; Gunsaldi et al., 2025; Holman et al., 2025). Gunsaldi et al. (2025) examined the impact of GenAI applications on student learning outcomes in science education, while Holman et al. (2025) systematically analyzed AI interventions in mathematics education. Purba et al. (2025) conducted an overview of AI-assisted learning, and Doo and Park (2026) confirmed the positive influence of ChatGPT on learning achievement through a meta-analysis. However, there has not been any SLR research that exclusively discusses the influence of GenAI on critical thinking skills in the context of science learning as a single focus of study, so this research offers an original contribution that is not yet available in the global literature (Zhu et al., 2026; Yusuf et al., 2024; Rahyuni et al., 2025; Hon, 2025; Gunsaldi et al., 2025).

METHODS

The method chosen in this study is the Systematic Literature Review (SLR). A Systematic Literature Review is a systematic and structured research method for collecting, assessing, and synthesizing all relevant literature on a specific topic (Kitchenham & Charters, 2007). This approach is explicit, transparent, and reproducible, minimizing selection bias and ensuring that all important sources are considered (Page et al., 2021; Fernandez & Dieguez, 2025). SLR differs from conventional narrative literature reviews by having a predefined protocol, a comprehensive search strategy, explicit inclusion and exclusion criteria, and a systematic assessment of methodological quality (Kitchenham & Charters, 2007; Braun & Clarke, 2006; Implom, 2025).

The reporting guidelines used in this study were PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), the latest international standard for reporting SLRs (Page et al., 2021). PRISMA 2020 updates previous guidelines by including broader database searches, more detailed risk-of-bias assessments, and more transparent reporting of the literature selection process (Page et al., 2021). These guidelines have been widely used in educational SLR research (Yulanda et al., 2026; Zhu et al., 2026; Su et al., 2026; Jia et al., 2025; Hon, 2025).

The initial steps involved determining the topic and formulating research questions. The researchers chose the topic of "the influence of Generative AI on critical thinking skills in science learning" as the main focus. Data for the literature study was obtained from three internationally reputable scientific databases: Scopus, ERIC (Education Resources Information Center), and Google Scholar. These three databases were selected based on their comprehensive coverage of education and science and their rigorous indexing standards (Yusuf et al., 2024; Zhu et al., 2026). The search string was constructed using the Boolean operators: ("generative AI" OR "ChatGPT" OR "large language model" OR "LLM") AND ("critical thinking" OR "higher-order thinking" OR "HOTS" OR "scientific reasoning") AND ("science education" OR "STEM education" OR "science learning"). The researchers limited the articles to 2020 to 2026 to ensure temporal relevance to the development of GenAI technology (Page et al., 2021; Kitchenham & Charters, 2007)

The inclusion criteria included: (1) articles published between 2020 and 2026; (2) examining the use of AI/GenAI in an educational context; (3) discussing critical thinking variables or HOTS; (4) focusing on the context of science or STEM education; (5) written in English; and (6) being peer-reviewed journal articles or proceedings (Kitchenham & Charters, 2007; Page et al., 2021; Fernandez & Dieguez, 2025). Exclusion criteria included: studies outside the educational context, opinion articles or editorials without empirical data, and duplicate articles from different databases. The data obtained were systematically compiled using thematic analysis (Braun & Clarke, 2006) with a deductive-inductive coding scheme that included categories of positive impacts, negative impacts, and moderating factors (Yulanda et al., 2026; Hon, 2025; An et al., 2026; Implom, 2025)

.

FINDINGS AND DISCUSSION

The results of this study were obtained through a literature review that summarized and analyzed various relevant sources discussing the influence of Generative AI on students' critical thinking skills in science learning. Research data, in the form of articles, were obtained through a screening process consisting of identification, screening, eligibility, and inclusion, in accordance with the PRISMA 2020 guidelines (Page et al., 2021; Kitchenham & Charters, 2007).

The identification stage was carried out by identifying articles from Scopus (n = 67), ERIC (n = 30), and Google Scholar (n = 93), resulting in a total of 190 identified records. The search was conducted using predefined search strings in these three databases (Yusuf et al., 2024; Zhu et al., 2026). At this stage, additional searches were conducted through reference lists of relevant articles to ensure no important literature was missed (Page et al., 2021).

The screening stage was carried out on the 190 identified articles. After removing 37 duplicates using a reference management tool, 153 articles remained for screening based on title and abstract by two independent reviewers. Of the 153 screened articles, 64 were excluded for not meeting the initial criteria, leaving 89 reports. Of these 89 reports, the full text of 8 reports was inaccessible, so 81 reports proceeded to the full-text screening stage (Page et al., 2021; Kitchenham & Charters, 2007; Fernandez & Dieguez, 2025).

The eligibility stage was conducted by reading the full text of the 81 reports. At this stage, 56 reports were excluded: 14 because the population was inappropriate, 15 because the intervention was inappropriate, 12 because the dependent variable was not critical thinking, 8 because data could not be extracted, and 7 were hidden duplicates (Page et al., 2021). The inclusion stage resulted in 25 studies that met all criteria and were included for in-depth analysis. These 25 studies come from three databases: 10 studies from Google Scholar, 7 studies from Scopus, and 8 studies from ERIC. The temporal distribution shows a significant increase in publications since 2023, reflecting the acceleration of research with the widespread adoption of GenAI tools (Jia et al., 2025; Hon, 2025; Zhu et al., 2026; An et al., 2026; Yusuf et al., 2024). A summary of the findings from the 25 included studies is systematically presented in Table 2 below

No.	Author(s) & Year	Journal	Research Findings
1	Implom, J. (2025)	*ICERI2025 Proceedings* DOI: 10.21125/iceri. 2025.2611	AI enhances inference, bias detection, and creativity when combined with domain-specific scaffolding and metacognitive strategies. Risks include surface-level learning, cognitive stagnation, and academic integrity concerns. LLMs act as catalysts for

			deeper cognitive engagement and adaptive reflective reasoning.
2	Zhu, Y. et al. (2026)	*European Journal of Education* DOI: 10.1111/ejed.70462	Generative AI provides significant benefits in STEM education. Major challenges include unreliable outputs, academic integrity concerns, cognitive dependency, reduced critical thinking, and accessibility barriers. Most studies focus on higher STEM education, while K–12 settings remain underrepresented.
3	Yulanda, N.I.P. et al. (2026)	*Journal of Learning Environments* DOI: 10.1080/23735082.2026.2672411	Meta-analysis of 17 studies (N = 2,052) found that inquiry-based AI produced a moderate-to-large positive effect on higher-order thinking skills (HOTS) (g = 0.696). ChatGPT (g = 0.754) outperformed traditional intelligent tutoring systems (g = 0.614). Problem-Based Learning (PBL) was the most effective approach (g = 0.812). Intervention duration significantly predicted effect size.
4	Fang, X. & Du, C. (2026)	*Journal of Educational Technology* (02 January 2026)	Generative AI promotes learning motivation, creativity, problem-solving, and metacognition. Students require higher-order skills (analysis, evaluation, and creation) to meaningfully adopt GenAI recommendations. AI literacy emerges as a key competency.
5	Su, C.S. et al. (2026)	*Journal of Computer Assisted Learning* DOI: 10.1002/jcal.70241	ChatGPT supports problem-solving (46.3%), instructional material design, and metacognitive scaffolding. Challenges include accuracy and reliability (46.3%), AI literacy (31.5%), and the need for strong teacher mediation (27.8%). Its value depends heavily on instructional design quality.
6	Rachmat, A. et al. (2025)	*IEEE EDUCON 2025* DOI: 10.1109/EDUCON62633.2025.11016417	Chatbot use did not significantly improve reflective thinking quantitatively; however, 21 of 29 respondents demonstrated qualitatively enhanced reflective thinking. Students also reported increased confidence in using AI as a learning tool after the intervention.
7	Jia, Y. et al. (2025)	*Springer Singapore* DOI: 10.1007/978-981-95-4499-8_32	A synthesis of 94 studies revealed that LLMs generally function as "Tutors and Assistants." Adoption models are evolving toward "Collaborative Partners" that foster HOTS. Students increasingly act as "Active Agents

			and Critical Evaluators," while educators become "Learning Architects."
8	Hon, K.K.L. (2025)	*Journal of Educational Technology Systems* DOI: 10.1177/00472395251400089	Evidence regarding GenAI's impact in higher education is mixed. Some studies report increased engagement, while others highlight excessive dependency. Effectiveness varies across disciplines, and more rigorous longitudinal research is needed.
9	An, Q. et al. (2026)	*Australasian Journal of Educational Technology* DOI: 10.14742/ajet.10561	A review of 39 articles identified seven GenAI usage patterns. Using GenAI as an evaluator and navigator yielded high success rates. When used as a creator, successes and challenges were balanced. Self-regulation applications showed the lowest success-to-challenge ratio.
10	Yusuf, A. et al. (2024)	*British Educational Research Journal* DOI: 10.1002/rev3.3489	Mapping of 407 publications identified eight discursive themes surrounding GenAI. Research gaps include limited K–12 studies, a lack of experimental research, and insufficient exploration of cultural dimensions in ethical concerns.
11	Chai, H. et al. (2026)	*Thinking Skills and Creativity* DOI: 10.1016/j.tsc.2026.10227	Meta-analysis of 32 studies showed a moderate positive correlation between GenAI use and student creativity ($g = 0.420$). Stronger effects were observed in natural sciences, collaborative learning environments, interventions lasting 1–4 weeks, and when GenAI was embedded within other platforms.
12	Oladokun, B.D. et al. (2026)	*The Journal of Academic Librarianship* DOI: 10.1016/j.acalib.2026.103253	Science and technology students demonstrated higher engagement with GenAI. Although GenAI improves accessibility, libraries remain valuable for ensuring information credibility. Universities should strengthen ethical, critical, and metacognitive AI literacy across curricula.
13	Wu, D. et al. (2024)	*Journal of Education for Library & Information Science* (Scopus)	Integrating reflective learning with GenAI enhances students' cognitive skills. GenAI assists in organizing ideas and providing alternative perspectives. Empirical evidence supports its role in inclusive learning environments grounded in diversity, equity, and inclusion (DEI) principles.

14	Ren, J. et al. (2026)	*Journal of Further and Higher Education*, 50(3) (Scopus)	SEM-PLS analysis of 482 students showed that GenAI adoption intentions are influenced by performance expectancy, self-efficacy, hedonic motivation, and habit. Educators should promote ethical use to preserve students' critical thinking and creativity.
15	Esh, M. & Ghosh, S. (2025)	*The Serials Librarian*, 86(3–4) DOI: 10.1080/0361526X.2025.2515173	Five unique research clusters were identified within AI and library studies. A weak correlation (r = 0.1049) between term popularity and citation frequency suggests that research quality and author influence are stronger determinants of impact.
16	Gorodetsky, V.I. (2024)	*Pattern Recognition and Image Analysis*, 34 DOI: 10.1134/S1054661824700184	Machine learning–based knowledge extraction methods are relatively mature. Expert knowledge remains underutilized despite its importance. Further research in knowledge engineering is required to process expert knowledge at scale.
17	He, S. & Lu, Y. (2024)	*Interactive Learning Environments*, 32(10) DOI: 10.1080/10494820.2024.2415444	In economic law courses, GenAI did not significantly improve cognitive levels from a knowledge-construction perspective. Behavioral patterns included autonomous-innovative (high-achieving), moderate (average), and less innovative (low-achieving) learners.
18	Williams, A. (2025)	*IJRES*, 11(3), 667–701 DOI: 10.46328/ijres.1302	LLMs facilitate personalized feedback and scientific terminology explanations that enhance understanding and critical thinking. Risks include plagiarism, algorithmic bias, and limited contextual understanding. LLMs enrich bioscience curricula through innovative teaching strategies.
19	Guo, Y. & Lee, D. (2023)	*Journal of Chemical Education*, 100 DOI: 10.1021/acs.jchemed.3c00599	ChatGPT-based activities in introductory chemistry increased students' confidence in asking deeper questions, analyzing information, and understanding complex concepts. ChatGPT challenged conventional thinking patterns and encouraged multiperspective reasoning.
20	Nawafleh, W. & Al-	*Educational Process: International*	An AI-based science module for seventh-grade students (N = 48) produced significant improvements in scientific concepts and

	Abbas, L. (2025)	*Journal*, 15 DOI: 10.22521/edupij.2025.15.92	critical thinking. Experimental group posttest scores (78.61) exceeded control group scores (60.27). AI enhanced interactivity, practical experiences, and motivation.
21	Lopes, L.A. (2024)	*Pedagogical Research*, 9(1) DOI: 10.29333/pr/14054	Creative challenge-based activities in remote natural science learning significantly increased student engagement and strengthened theory–practice connections. The approach stimulated creativity, critical thinking, and essential skills development.
22	Atmojo, I.R.W. et al. (2025)	*Educational Process: International Journal*, 15 DOI: 10.22521/edupij.2025.15.131	AR-assisted Project-Based Learning (PjBL) significantly improved sixth-grade students' critical thinking (experimental: 79.61; control: 60.27). AR and PjBL increased motivation, engagement, and critical thinking. The FRISCO framework was used to assess critical thinking indicators.
23	Fadillah, M.A. et al. (2024)	*Journal of Baltic Science Education*, 23(6) DOI: 10.33225/jbse/24.23.1178	ChatGPT dimensions of Cognitive Quality (CQ), Meaningful Engagement (ME), and Adaptive Thinking (AT) were significantly correlated with Indonesian high school students' HOTS (N = 334). CQ showed the strongest correlation. ChatGPT demonstrates potential to support inquiry-based physics learning and HOTS development.
24	Uyar, A. (2025)	*Journal of Baltic Science Education*, 24(3) DOI: 10.33225/jbse/25.24.552	Turkish STEM educators (N = 15) reported that AI saves time, personalizes content, and increases motivation. However, excessive use may hinder problem-solving and critical thinking. Technical infrastructure remains a major challenge.
25	Kafilongo, E. et al. (2026)	*Educational Process: International Journal*, 22 DOI: 10.22521/edupij.2026.22.39	Integrating AI into student-centered learning requires a gradual approach, culturally responsive AI tools, and locally relevant datasets. Teacher–AI–student collaboration enhances learner autonomy, real-time feedback, and differentiated instruction.

Discussion

Based on the description in Table 2, all 25 articles reviewed discussed the impact of the use of Generative AI in educational contexts on students' critical thinking and higher-order cognitive abilities. Critical thinking, as a fundamental competency in science learning, requires students not only to understand concepts but also to analyze, evaluate, and apply scientific knowledge in various real-life contexts (Facione, 1990; Anderson & Krathwohl, 2001; Paul & Elder, 2006). The GenAI tools used in various studies vary widely: from ChatGPT (Guo & Lee, 2023; Fadillah et al., 2024; Su et al., 2026), LLM systems integrated into learning platforms (Rachmat et al., 2025; Wu et al., 2024; Jia et al., 2025), to AI-based instructional modules (Nawafleh & Al-Abbas, 2025; Kafilongo et al., 2026). This diversity indicates that GenAI's impact cannot be understood in a generalized manner but must be analyzed within the context of its intended use (Yusuf et al., 2024; An et al., 2026; Hon, 2025).

The Positive Impact of GenAI on Critical Thinking Skills in Science Learning

Empirical evidence from various included studies consistently shows that the use of GenAI integrated with inquiry-based and problem-solving approaches has a significant positive impact on students' critical thinking skills and HOTS. Yulanda et al. (2026) conducted a meta-analysis of 17 studies (N = 2,052) and found that AI combined with inquiry-based learning produced a medium to large effect size (g = 0.696, 95% CI [0.607, 0.785]). ChatGPT produced a larger effect size (g = 0.754) than a traditional intelligent tutoring system (g = 0.614), and problem-based learning demonstrated the highest effectiveness (g = 0.812), confirming that the quality of instructional design is a key variable (Su et al., 2026; Implom, 2025; Uyar, 2025; Nawafleh & Al-Abbas, 2025; Guo & Lee, 2023).

Nawafleh and Al-Abbas's (2025) experimental study with seventh-grade students (N = 48) provided strong causal evidence: an AI-based science learning module produced a statistically significant difference in scientific concept acquisition and critical thinking between the experimental group (posttest mean 78.61) and the control group (60.27). Similarly, Atmojo et al. (2025) found that an Augmented Reality-assisted PjBL model significantly improved the critical thinking of sixth-grade students, with the experimental class achieving a posttest score of 79.61 compared to the control class's score of 60.27. Both studies employed a pre-posttest design with a control group, providing a strong comparative basis for identifying the causal effects of AI-based technology on critical thinking (Fadillah et al., 2024; Guo & Lee, 2023; Williams, 2025; Chai et al., 2026; Lopes, 2024).

From a Vygotskyite constructivist perspective, the mechanism of this positive impact can be explained through the concept of scaffolding within the Zone of Proximal Development (ZPD): GenAI functions as a mediating tool that helps students achieve understandings they cannot yet achieve independently. Jia et al.'s (2025) research, which synthesized 94 empirical studies, identified a shift in the role of LLM from mere 'tutors and assistants' to 'collaborative partners' that encourage co-creation and HOTS. This shift shifts students' roles from passive recipients to active agents and critical evaluators—a transformation that reflects the development of critical thinking (Jia et al., 2025; Fang & Du, 2026; An et al., 2026; Wu et al., 2024; Rachmat et al., 2025).

Guo and Lee (2023) provide strong contextual evidence from college-level science education: ChatGPT-based activities in an introductory chemistry course significantly increased students' confidence in asking in-depth questions, analyzing information, and understanding complex concepts. Students explicitly reported that ChatGPT provided diverse perspectives and challenged their existing ways of thinking. This aligns with the findings of Fadillah et al. (2024) who found that the quality dimensions of ChatGPT were significantly correlated with HOTS of Indonesian high school students, and were predictors of effective physics inquiry learning (Fadillah et al., 2024; Williams, 2025; Lopes, 2024; Nawafleh & Al-Abbas, 2025; Atmojo et al., 2025).

Chai et al.'s (2026) meta-analysis of 32 studies provides additional relevant perspective: GenAI-based learning showed a moderate positive correlation with student creativity (g = 0.420), with the largest effects found in the natural sciences and collaborative learning. Creativity and critical thinking are two interrelated competencies within the scientific reasoning framework. Research by Wu et al. (2024) also showed that integrating reflective learning with GenAI improved students' cognitive skills, with AI helping to organize ideas and offer alternative perspectives. Oladokun et al. (2026) added that students in science and technology fields showed higher engagement with GenAI than in other disciplines (Chai et al., 2026; Lopes, 2024; Wu et al., 2024; Atmojo et al., 2025; Oladokun et al., 2026).

Negative Impact of GenAI on Critical Thinking Skills

On the other hand, various studies have identified potential negative impacts that require serious attention in the context of science education. Zhu et al. (2026) in a systematic synthesis of 55 STEM studies found that GenAI poses challenges for learners, including: unreliable and inaccurate output, academic integrity concerns, cognitive dependency, reduced critical thinking, and digital accessibility barriers. Uyar (2025) in a phenomenological study of 15 Turkish STEM educators explicitly found that excessive or unconsidered use of AI can hinder students' problem-solving and critical thinking skills. Implom (2025) identified three main risks documented in 44 articles: surface learning, cognitive stagnation, and challenges to academic integrity (Zhu et al., 2026; Uyar, 2025; Hon, 2025; Implom, 2025; He & Lu, 2024).

Rachmat et al.'s (2025) study yielded an important nuanced finding: quantitatively, the use of chatbots did not result in significant improvements in students' reflective thinking skills. Qualitatively, however, 21 of the 29 respondents showed evidence of increased reflective thinking in their open-ended reports. This paradox suggests that AI's impact on critical thinking may be more process-driven than product-driven—a critical methodological gap in the literature that relies on standardized tests (Rachmat et al., 2025; Hon, 2025; An et al., 2026; Yusuf et al., 2024; He & Lu, 2024).

An et al. (2026) specifically identified that when students used GenAI as a 'creator'—delegating the creation of original content to the AI—the proportion of successes and challenges was nearly equal. More concerning, using GenAI as a self-regulation supporter resulted in the lowest success-to-challenge ratio among all use cases. He and Lu (2024) reinforced this concern by finding that GenAI did not significantly improve cognitive levels from a knowledge construction perspective, with low-achieving students exhibiting 'less innovative' behavioral patterns in their interactions with AI (He & Lu, 2024; An et al., 2026; Ren et al., 2026; Hon, 2025; Uyar, 2025).

The dimensions of algorithmic bias and AI hallucinations pose a particularly dangerous threat in the context of science learning, where the validity and accuracy of information are fundamental epistemological standards. Williams (2025) identified that key concerns in bioscience education include plagiarism, algorithmic bias, and limited contextual understanding of LLMs. Su et al. (2026) found accuracy and reliability to be the greatest challenges (46.3% of studies), requiring explicit verification protocols. When students with low AI literacy receive GenAI output without critical verification, they risk constructing erroneous scientific understandings (Williams, 2025; Su et al., 2026; Implom, 2025; Zhu et al., 2026; Fang & Du, 2026).

Factors Influencing the Impact of GenAI on Critical Thinking

A synthesis across 25 studies consistently identified that the impact of GenAI on critical thinking is mediated by a series of contextual factors that are more determinant than the type of AI tool used itself. The first and most consistent factor is students' AI literacy. Fang and Du (2026) asserted that students require higher-order thinking skills analysis, evaluation, and creation to meaningfully adopt GenAI recommendations. Ren et al. (2026) found, through SEM-PLS analysis of 482 university students, that self-efficacy in using GenAI was a significant predictor of adoption intention and actual use. Oladokun et al. (2026) recommended that universities strengthen AI literacy by encouraging ethical, critical, and metacognitive engagement (Fang & Du, 2026; Ren et al., 2026; Oladokun et al., 2026; Implom, 2025; Su et al., 2026).

The second factor is the teacher's instructional design and pedagogical strategies. Su et al. (2026) explicitly concluded that the educational value of ChatGPT depends on the quality of the instructional design, not the capabilities of the tool itself. Jia et al. (2025) identified that shifting the role of educators to 'Learning Architects' is key to optimizing the impact of AI on HOTS. Uyar (2025) found that although AI has been shown to support cognitive development, its use without careful pedagogical consideration actually hinders problem-solving and critical thinking. Kafilongo et al. (2026) added that AI integration requires a phased approach with relevant local datasets and AI tools that are responsive to the local cultural context (Su et al., 2026; Jia et al., 2025; Uyar, 2025; Rachmat et al., 2025; Kafilongo et al., 2026).

The third factor is the duration and intensity of the intervention. Yulanda et al. (2026) identified intervention duration as a significant predictor of effect size in a meta-analysis, with 1-4 weeks of interventions proving optimal for developing creativity and HOTS. Chai et al. (2026) confirmed these findings in the context of creativity: 1-4 weeks of interventions produced greater effects than shorter or longer interventions. Williams (2025) highlighted that LLM only effectively facilitates critical thinking when used in a structured and ongoing context, not in ad-hoc access without a pedagogical framework (Yulanda et al., 2026; Chai et al., 2026; Williams, 2025; Nawafleh & Al-Abbas, 2025; Fadillah et al., 2024).

The fourth factor is task and assessment design. An et al. (2026) found that how students use GenAI as a translator, navigator, evaluator, or creator significantly determines the learning success rate. Lopes (2024) demonstrated that a creative challenge-based approach to science learning simultaneously stimulates critical thinking and creativity. He and Lu (2024) identified that learning behavior patterns differ significantly based on student achievement levels, indicating that tasks that explicitly demand cognitive autonomy are more effective for high-achieving students (An et al., 2026; Lopes, 2024; Atmojo et al., 2025; He & Lu, 2024; Kafilongo et al., 2026).

Based on a synthesis of these four factors, a consistent pattern emerged across the 25 articles reviewed: the impact of GenAI on students' critical thinking is conditional, not universal (Jia et al., 2025; Zhu et al., 2026; Hon, 2025). GenAI has the potential to be a cognitive amplifier when used in a structured learning context with appropriate scaffolding, adequate AI literacy, and task design that challenges critical thinking (Yulanda et al., 2026; Nawafleh & Al-Abbas, 2025; Guo & Lee, 2023; Fadillah et al., 2024). Conversely, without clear pedagogical guidance, GenAI risks becoming a cognitive substitute that replaces, rather than supports, students' scientific thinking processes, as evidenced by He and Lu (2024), An et al. (2026), and Rachmat et al. (2025). These findings align with Vygotsky's constructivist theory: AI is most effective when it functions as a scaffold that encourages the internalization of cognitive processes (Jia et al., 2025; Fang & Du, 2026; Wu et al., 2024; Kafilongo et al., 2026; Uyar, 2025).

CONCLUSION

Based on the results and discussion described above, it can be concluded that the impact of GenAI on students' critical thinking skills in science learning is ambivalent, conditional, and highly dependent on the context of its use. Of the 25 studies included through the PRISMA 2020 protocol, there is consistent evidence that GenAI integrated within an inquiry-based and problem-solving pedagogical framework results in significant improvements in HOTS and critical thinking, as demonstrated by Yulanda et al. (2026) with a g effect of 0.696, Nawafleh and Al-Abbas (2025) with a posttest difference of 18.34 points, and Fadillah et al. (2024) through a significant ChatGPT-HOTS correlation in the Indonesian context. Conversely, the use of GenAI without adequate pedagogical scaffolding has the potential to lead to cognitive dependency, surface learning, and a decline in independent reasoning abilities as identified by Zhu et al. (2026), Implom (2025), and Uyar (2025). The four main moderating factors consistently found were: students' AI literacy, the quality of teachers' instructional design, the duration of the intervention, and the structure of tasks and assessments.

REFERENCES

An, Q., Koh, J. H. L., & Liu, Q. (2026). Generative Artificial Intelligence in higher education: A systematic review of student use and learning outcomes. Australasian Journal of Educational Technology. https://doi.org/10.14742/ajet.10561

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Atmojo, I. R. W., Saputri, D. Y., Dewi, R. K., Salimi, M., Roslan, R. M., & Halim, L. (2025). The effect of the implementation of a project-based learning model assisted by augmented reality on sixth graders' critical thinking skills on solar system materials. Educational Process: International Journal, 15, e2025131. https://doi.org/10.22521/edupij.2025.15.131

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Chai, H., Guo, F., Wang, J., Song, X., & Hu, T. (2026). Uncovering the effect of Gen AI-assisted learning on students' creativity: A meta-analysis from quasi-experimental evidence. Thinking Skills and Creativity, 62(A), 102227. https://doi.org/10.1016/j.tsc.2026.10227

Doo, M. Y., & Park, Y. (2026). A meta-analysis of ChatGPT's influence on learning achievement. [Google Scholar].

Esh, M., & Ghosh, S. (2025). Analysing thematic structures and citation dynamics in artificial intelligence (AI) and library research: A bibliometric and machine learning approach. The Serials Librarian, 86(3-4), 161-177. https://doi.org/10.1080/0361526X.2025.2515173

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. The California Academic Press.

Fadillah, M. A., Usmeldi, & Asrizal. (2024). The role of ChatGPT and higher-order thinking skills as predictors of physics inquiry. Journal of Baltic Science Education, 23(6), 1178-1192. https://doi.org/10.33225/jbse/24.23.1178

Fang, X., & Du, C. (2026). A systematic review of empirical studies fostering generative AI literacy in primary and secondary education. Journal of Educational Technology (published online 02 January 2026).

Fernandez, M., & Irimia-Dieguez, A. (2025). Links between food tourism and the local economy: A systematic literature review. Sustainable Futures, 100959.

Gorodetsky, V. I. (2024). From knowledge engineering to knowledge science. Pattern Recognition and Image Analysis, 34, 440-447. https://doi.org/10.1134/S1054661824700184

Gunsaldi, M. S., Guner, E. G., Uckan, M., & Bati, K. (2025). The impact of generative AI applications on student learning outcomes in science education: A systematic review. [Google Scholar].

Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. Journal of Chemical Education, 100, 4876-4883. https://doi.org/10.1021/acs.jchemed.3c00599

He, S., & Lu, Y. (2024). The effectiveness of Gen AI in assisting students' knowledge construction in humanities and social sciences courses: Learning behaviour analysis. Interactive Learning Environments, 32(10), 7041-7062. https://doi.org/10.1080/10494820.2024.2415444

Holman, K., Marino, M., Vasquez, T., Taub, M., Hunt, J., & Tazi, Y. (2025). Artificial intelligence interventions in mathematics education: A systematic literature review. [Google Scholar].

Hon, K. K. L. (2025). Generative AI in higher education: A systematic review of its effects on learning outcomes and academic performance. Journal of Educational Technology Systems, 54(3). https://doi.org/10.1177/00472395251400089

Implom, J. (2025). Artificial intelligence and critical thinking in higher education: A scoping review. ICERI2025 Proceedings, 9253-9263. ISBN: 978-84-09-78706-7. https://doi.org/10.21125/iceri.2025.2611

Jia, Y., Li, C., Chan, A., & Li, Q. (2025). Human-AI collaboration in the STEM classroom: A systematic literature review of GenAI as a complement in higher education. In Proceedings of International Conference. Springer, Singapore. https://doi.org/10.1007/978-981-95-4499-8_32

Kafilongo, E., et al. (2026). Integrating AI in student-centered learning: Phased approaches and cultural responsiveness. Educational Process: International Journal, 22, e2026039. https://doi.org/10.22521/edupij.2026.22.39

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, Keele University.

Lopes, L. A. (2024). Creative challenge to stimulate student engagement in natural science education in distance learning. Pedagogical Research, 9(1), em0184. https://doi.org/10.29333/pr/14054

Majeed, A., et al. (2025). The dual impact of artificial intelligence on critical thinking. [Lengkapi dengan jurnal dan DOI sesuai referensi asli].

Nawafleh, W., & Al-Abbas, L. (2025). Developing an artificial intelligence-based science instructional module and measuring its effect on acquiring scientific concepts and critical thinking skills among seventh-grade female students. Educational Process: International Journal, 15, e2025092. https://doi.org/10.22521/edupij.2025.15.92

Oladokun, B. D., Atuase, D., Tachie-Donkor, G., Mosweu, T. L., & Gaitanou, P. (2026). University libraries in the age of generative AI: Perceptions and usage among students in developing countries. The Journal of Academic Librarianship, 52(4), 103253. https://doi.org/10.1016/j.acalib.2026.103253

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Paul, R., & Elder, L. (2006). Critical thinking: Tools for taking charge of your learning and your life (2nd ed.). Pearson Prentice Hall.

Pham, T., et al. (2025). Personalized learning through artificial intelligence: Evidence from science education contexts. [Lengkapi dengan jurnal dan DOI sesuai referensi asli].

Purba, S. W. D., Silitonga, B. N., & Yang, J. J. (2025). AI-assisted learning: A systematic review. [Google Scholar].

Rachmat, A., Watterson, C., & Lundqvist, K. (2025). The impact of chatbots on students' reflective thinking in introductory programming course. In Proceedings of IEEE EDUCON 2025. https://doi.org/10.1109/EDUCON62633.2025.11016417

Rahyuni, M., Ashadi, Triastuti, A., Hidayati, S., Salido, A., Ero, P. E. L., Marlini, C., Zefrin, & Al Fuad, Z. (2025). Critical thinking in the age of AI: A systematic review of AI's effects on higher education. [Google Scholar].

Ren, J., Shahzad, M. F., Abbas, J., Al-Sulaiti, K. I., & Pilar, L. (2026). Factors influencing generative artificial intelligence such as ChatGPT adoption and actual use in higher education: An extended UTAUT2 perspective. Journal of Further and Higher Education, 50(3). [Scopus].

Simoni, A., et al. (2025). Critical thinking in the age of generative AI: Research directions and frameworks. [Lengkapi dengan jurnal dan DOI sesuai referensi asli].

Su, C. S., Hsu, C. C., & Gonzalez Campos, J. A. (2026). ChatGPT as a resource to strengthen mathematics teaching and learning: A systematic review. Journal of Computer Assisted Learning. https://doi.org/10.1002/jcal.70241

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive

Science, 12(2), 257-285.

Uyar, A. (2025). Effective use of artificial intelligence tools in STEM education: Perspectives of STEM educators. Journal of Baltic Science Education, 24(3). https://doi.org/10.33225/jbse/25.24.552

Williams, A. (2025). Critical evaluation of the potential of large language models in bioscience higher education: A conversation with ChatGPT. International Journal of Research in Education and Science (IJRES), 11(3), 667-701. https://doi.org/10.46328/ijres.1302

Wu, D., Wang, H.-T., & DelaCruz, J. J. (2024). Reflective learning with AI assistance in information science classrooms for DEI. Journal of Education for Library and Information Science. [Scopus].

Yulanda, N. I. P., Gestiardi, R., Subanji, Nusantara, T., & Setyosari, P. (2026). Artificial intelligence-enhanced inquiry-based learning and higher-order thinking skills: A meta-analysis. (Published online 22 May 2026). https://doi.org/10.1080/23735082.2026.2672411

Yusuf, A., Pervin, N., Roman-Gonzalez, M., & Noor, N. M. (2024). Generative AI in education and research: A systematic mapping review. Review of Education, 12(3). https://doi.org/10.1002/rev3.3489

Zhu, Y., Jiang, H., & Zhong, L. (2026). Generative Artificial Intelligence-assisted STEM education: A systematic review and future research agenda. European Journal of Education. https://doi.org/10.1111/ejed.70462

.

.