Momentum: Physics Education Journal, 10(1), 2026

Investigating the influence of generative ai integration within Problem Based Learning on students' critical thinking in introductory physics courses

Putu Prima Juniartina[1]*, Ni Luh Putu Mery Marlinda[1], I Made Oka Riawan[2], Kadek Dwi Hendratma Gunawan[3]

[1]Department of Science Education, Universitas Pendidikan Ganesha, Udayana Street No.1, Bali, 81116, Indonesia

[2]Department of Biology, Universitas Pendidikan Ganesha, Udayana Street No.11, Bali, 81116, Indonesia

[3]Department of Science Education, Universitas Sebelas Maret, Ir. Sutami Street, No.36A Surakarta, 57126, Indonesia

*Corresponding author, email: prima.juniartina@undiksha.ac.id

Article History
Received: 26 December 2025
Revised: 17 April 2026
Accepted: 28 May 2026

Keywords
Critical thinking
Generative AI
Physics education

Abstract

While the proliferation of Generative Artificial Intelligence (GenAI) offers transformative potential in higher education, its specific impact on critical thinking when integrated into structured pedagogical frameworks remains underexplored. This study investigates the integration of GenAI as a cognitive scaffold within a Problem-Based Learning (PBL) framework and its effect on the critical thinking skills of undergraduate students in an Introductory Physics course. Employing a quantitative quasi-experimental approach with a nonequivalent pretest-posttest control group design, the research involved 38 Science Education students. Participants were divided into an experimental group utilizing GenAI-assisted PBL and a control group receiving conventional instruction. Data were collected using 15 essay items assessing critical thinking based on Ennis's taxonomy and analyzed via ANOVA and Normalized Gain (N-Gain). Results revealed a statistically significant difference (F=100.07; p<0.001), with the experimental group achieving a "Medium" gain (N-Gain = 0.477) compared to the control group's "Low" gain (N-Gain = 0.189). These findings address a critical gap by demonstrating that when GenAI is explicitly paired with the evaluative demands of PBL, it functions effectively as intelligent scaffolding rather than causing cognitive offloading, underscoring the necessity of embedding AI literacy into the modern science curriculum.

Juniartina, P. P., Marlinda, N. L. P. M., Riawan, I. M. O., & Gunawan, K. D. H. (2026). Investigating the influence of generative ai integration within Problem Based Learning on students' critical thinking in introductory physics courses. *Momentum: Physics Education Journal*, *10*(1), 37-46. https://doi.org/10.21067/mpej.v10i1.13647

1. Introduction

The contemporary global higher education landscape is experiencing a seismic disruption, catalyzed by rapid advancements in Generative Artificial Intelligence (GenAI). Technologies powered by Large Language Models (LLMs) such as OpenAI's ChatGPT, Google's Gemini, and recent innovations like DeepSeek have fundamentally altered the paradigms of information access, synthesis, and production (Kasneci et al., 2023a). In the academic ecosystem, GenAI has transcended its role as a passive tool for information retrieval, evolving into an external cognitive entity capable of performing logical reasoning, solving complex mathematical problems, and writing programming code (Helal et al., 2025). This phenomenon presents unprecedented challenges and opportunities for science educators, particularly within Introductory Physics, a discipline requiring analytical rigor and profound conceptual understanding (Nemalynne et al., 2023).

Introductory Physics, Basic Physics, functioning as a gateway course for science and engineering majors, has traditionally acted as a rigorous academic filter (Şengül, 2023). Students typically struggle not with the rote memorization of formulas, but with the capacity to bridge abstract concepts with tangible physical phenomena and apply these principles to novel problem-solving scenarios key facets of critical thinking (Alarbi et al., 2024). Conventional instructional methods frequently fail to bridge this cognitive gap due to the limitations on lecturers' ability to provide personalized feedback (Bogdanova & Snoeck, 2019). In this context, GenAI offers transformative

doi: 10.21067/mpej.v10i1.13647

ISSN: 2548-9135

potential as a "virtual tutor," delivering personalized explanations and real-time instructional scaffolding (Younis, 2025). Nevertheless, the integration of GenAI into physics education is not devoid of controversy. A primary concern in current academic discourse is the risk of cognitive offloading (Gerlich, 2025a). where an over-reliance on machine reasoning could erode students' intrinsic critical thinking capabilities.

The evolution of literature on GenAI in science education from 2020 to 2025 reveals a trajectory from skepticism to constructive pedagogical exploration. While early studies focused on plagiarism and factual inaccuracies (Khan & Saunderson, 2024). Research published in 2024 and 2025 has begun to characterize GenAI as a co-learning agent (Alfarwan, 2025). For example, Mahligawati et al. (2023) demonstrated that physics curricula integrating AI-assisted experimental data analysis significantly aided students in addressing complex problems. Similarly, Kotsis (2025) compared ChatGPT and DeepSeek, concluding that the synergy between ChatGPT's interactivity and DeepSeek's reasoning depth creates a highly adaptive science learning environment. Given GenAI's susceptibility to "hallucinations" generating plausible but erroneous information the ability to verify, critique, and evaluate AI outputs has emerged as a new manifestation of critical thinking (Jho, 2024). This study adopts Ennis's taxonomy for the operational definition of critical thinking, encompassing the skills of problem formulation, argumentation, deduction, induction, and evaluation (Worthington, 2019).

The Problem Based Learning (PBL) model was selected as the instructional framework due to its high compatibility with critical thinking development (Harahap et al., 2025). Research by Ayu et al (2024) and other systematic reviews confirm PBL as an effective strategy for enhancing physics conceptual understanding; the integration of AI is hypothesized to amplify these effects. Despite the growing body of literature acknowledging GenAI as a co-learning agent, a significant research gap remains regarding how to operationalize this technology within specific active learning environments to prevent cognitive offloading. Previous studies have primarily focused on unstructured AI usage, leaving its potential as a systematic pedagogical tool underexplored. This study addresses this gap by synthesizing GenAI with the Problem-Based Learning (PBL) framework. The relationship between PBL and GenAI is highly complementary: while GenAI serves as an immediate, on-demand cognitive scaffold to retrieve data and clarify basic concepts, the inherently ill-structured problems in PBL compel students to critically evaluate, cross-reference, and synthesize the AI's output. Therefore, the PBL framework transforms the student's role from a passive consumer of AI-generated answers into an active validator. Consequently, this study aims to investigate. Consequently, this study aims to investigate the impact of Generative AI usage on the critical thinking skills of undergraduate students in the Science Education at Universitas Pendidikan Ganesha (UNDIKSHA) in the Introductory Physics course. Specifically, it seeks to determine whether structured, pedagogically grounded GenAI integration (within a PBL framework) produces superior critical thinking outcomes compared to conventional instruction.

2. Method

Conducted at the Undergraduate Science Education Program of Universitas Pendidikan Ganesha, this study utilized a quantitative approach based on a quasi-experimental method. The specific framework applied was the nonequivalent pretest-posttest control group design, as depicted in Figure 1.

E O_1 X_1 O_2

\- -

K O_1 X_2 O_2

Figure 1. Non-Equivalent Pretest Posttest Control Group Design Experimental Design

Description:
E1 = Experimental group, students followed the problem-based learning model with Generative AI
K = Control class, students follow the conventional learning model
X1 = Treatment with a problem-based learning model with Generative AI
X2 = Treatment with conventional learning models
O1 = pre-test
O2 = post test

Given that the entire population was included in the study, a total sampling (census) technique was utilized. The sample comprised 38 undergraduate students from the Science Education program at Undiksha, distributed into two distinct groups: an experimental class and a control class. Data regarding critical thinking skills were collected using an assessment instrument consisting of 15 essay items. To establish the instrument's rigor, it was subjected to a content validity evaluation by two experts in Physics. The specific indicators for critical thinking skills are outlined in Table 1.

Table 1. Indicators for Critical Thinking Skills

No	Indicators for critical thinking skills	Competency Description
1	Formulating the Problem	Formulate questions that provide direction to obtain answers; identify the core of the physics problem.
2	Providing Arguments	Provide appropriate reasons; show differences and similarities; construct a complete logical argument.
3	Making Deductions	Deducing logical consequences from general principles (e.g., Newton's Laws); interpreting logical conditions.
4	Performing Induction	Conducting investigations/data collection; drawing general conclusions/hypotheses from specific data; providing logical assumptions.
5	Conducting an Evaluation	Assess the credibility of statements or solutions based on facts, principles, or guidelines; offer alternative solutions.
6	Decide and implement	Selecting the best solution from various possibilities and determining the execution strategy.

(Ennis, 1985)

Critical thinking proficiency was assessed using the normalized gain (N-Gain), which represents the normalized differential between pretest and posttest scores. This metric was subsequently utilized in hypothesis testing via Analysis of Variance (ANOVA). The N-Gain metric is advantageous as it allows for the distinct characterization of each unit of analysis, differentiating performance even when raw score differences appear identical. N-Gain values range from -1 to 1; a positive value reflects an improvement in learning outcomes post-instruction, while a negative value indicates a decline. The calculation for the N-Gain score is defined in Equation 1.

$$\text{NGain} = \frac{Score\ posttest - Score\ pretest}{Score\ Ideal - Score\ pretest} \tag{1}$$

To see the category of the magnitude of the increase in the N-Gain score, you can refer to the normalized Gain criteria in Table 2.

Table 2. Normalized Gain Criteria

No	N-Gain Value	Interpretation
1	$0.70 \le g \le 1.00$	High
2	$0.30 \le g < 0.70$	Medium
3	$0.00 < g < 0.30$	Low
4	$g = 0{,}00$	There was no increase
5	$-1.00 \le g < 0.00$	There was a decline

(Hake, 1998)

The criterion for statistical significance was established by comparing the calculated F-value against the tabular F-value at a 5% significance level. All computational analyses were facilitated by SPSS for Windows. Before proceeding with hypothesis testing, the data underwent prerequisite assumption testing to ensure its suitability for ANOVA.

3. Results and Discussion

3.1. Results Study

Empirical data derived from 38 participants underscored distinct differences in the learning trajectories of the experimental and control groups. Descriptive analysis of the raw scores offered initial insights into the intervention's efficacy. Pre-intervention assessments indicated that both groups possessed homogeneous baseline capabilities, thereby mitigating concerns regarding selection bias. Conversely, a pronounced divergence in final achievements was observed upon the conclusion of the intervention. Table 3 summarizes the descriptive statistics for critical thinking scores. Notably, the experimental group utilizing GenAI as a co-learning agent exhibited a substantial enhancement in posttest scores relative to the control group. Next, comparison of average pretest and posttest scores can be seen in Figure 2.

Table 3. Comparison of Average Pretest and Posttest Scores

Group	Number of Samples (N)	Average pre-test Score	Average Post-test Score	Difference (Gross Gain)
Experiment (GenAI)	19	48.37	72.68	+24.31
Control (Conventional)	19	48.00	57.74	+9.74

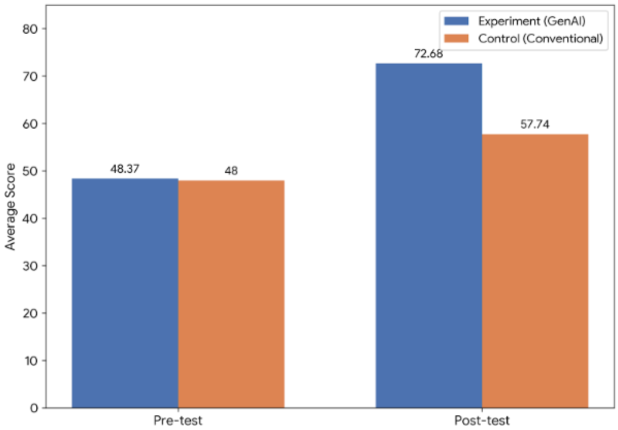

Figure 2. Comparison of Average Pretest and Posttest Scores

As illustrated in Table 3 and visually represented in Figure 2, both groups demonstrated homogeneous baseline capabilities during the pretest. However, a pronounced divergence is evident in the posttest results, visually highlighting the significant efficacy of the GenAI intervention compared to conventional instruction. These findings suggest a strong initial indication of the intervention's efficacy.

To ensure a more rigorous evaluation of instructional efficacy, Normalized Gain (N-Gain) scores were computed for each participant. As shown in Table 4, the N-Gain analysis was employed to mitigate the ceiling effect and establish a standardized metric of academic growth. Collectively, these data demonstrate the intervention's impact relative to the maximum potential gain available to each student.

Table 4. N-Gain Score Analysis and Categorization

Group	Average N-Gain	Standard Deviation	Category (Hake)
Experiment (GenAI)	0.477	0.108	Medium
Control (Conventional)	0.189	0.064	Low

The data presented in Table 4 reinforces the preceding descriptive findings. The experimental group successfully attained the 'Medium' gain category (0.477) a noteworthy accomplishment within the demanding context of physics instruction. Conversely, the control group was confined to the 'Low' category (0.189). It is crucial to observe that the standard deviation for the experimental group (0.108) exceeded that of the control group (0.064), indicating a higher degree of heterogeneity in the impact of AI adoption. While some students likely maximized AI as instructional scaffolding (approaching an N-Gain of 0.7), others may have encountered challenges in integrating the technology; this phenomenon will be elaborated upon in the Discussion section.

3.2. Hypothesis Testing Results

Before proceeding to hypothesis testing, the suitability of the data for testing with ANOVA is verified through prerequisite tests:

3.2.1. Normality Testing Results

The normality test was used to check whether the N-Gain data from each group was normally distributed. The normality assumption was met if the p-value was greater than 0.05. Normality testing results can be seen in Table 5.

Table 5. Normality Testing Results

Group	N	Statistic Kolmogorov Smirnov	Sig. (P)
Experiment	19	0.9617	0.6058
Control	19	0.9796	0.9375

Based on Table 5, the p-value for both groups (0.6058 and 0.9375) are greater than 0.05, so it can be concluded that the N-Gain data in both groups are normally distributed.

3.2.2. Homogeneity Testing Results

This test is conducted to determine whether the variance (diversity) of the two groups is the same or homogeneous. The assumption of homogeneity is met if the p-value is greater than 0.05. Homogeneity testing results can be seen in Table 6.

Table 6. Homogeneity Testing Results

Variable		Levene Statistic	df1	df2	Sig (p)
N-Gain	Based On Mean	0.6376	1	36	0.4298
	Based On Median	0.831	1	36	0.364
	Based On Median and with adjusted df	0.831	1	36	0.364
	Based on trimmed mean	1.138	1	36	0.288

As indicated in Table 6, the p-value exceeds 0.05, confirming that the variances across both groups are homogeneous. Having satisfied these prerequisite assumptions, the data qualify for subsequent analysis via parametric testing specifically ANOVA to identify any significant disparities between the groups

3.2.3. ANOVA Testing Results

Following the fulfillment of these assumptions, a One Way ANOVA was conducted to evaluate the research hypothesis regarding significant disparities in critical thinking capabilities between the GenAI-assisted group and the non-GenAI group. Statistical significance was determined by comparing the calculated F-value against the critical table F-value at a significance level of 5%. Conventionally, a difference is deemed significant when the p-value falls below 0.05. Table 7 shown a result of the ANOVA test.

Table 7. Results of the ANOVA Test

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-value	p-value
Intergroup	0.7864	1	0.7864	100.0738	0.0000
in Group	0.2829	36	0.0079	-	-
Total	1.0692	37	-	-	-

As presented in Table 7, the ANOVA yielded a calculated F-value of 100.0738 and a p-value of 0.0000. Given that the p-value (0.0000) falls below the 0.05 threshold, a statistically significant difference exists in the critical thinking capabilities of the experimental versus the control group. Consequently, these findings demonstrate that integrating GenAI into introductory Physics education significantly enhances students' critical thinking proficiency.

3.3. Discussion

These research findings provide compelling empirical validation of Generative AI's potential to serve as a catalyst for critical thinking in physics education provided, crucially, that it is deployed within a structured pedagogical framework. Nevertheless, these statistical metrics merely scratch the surface of a far more complex phenomenon. Why does GenAI foster critical thinking in physics, contradicting prevalent fears that it fosters cognitive lethargy among students? Furthermore, notwithstanding the significant gains observed, why did the overall achievement plateau within the 'Medium' category?.

3.3.1. GenAI as Intelligent Scaffolding

The experimental group's advantage can be attributed to the mechanism of cognitive scaffolding. In conventional physics settings, students frequently grapple with cognitive overload (Wu & Valente, 2020), as they must simultaneously recall formulas, visualize force systems, execute algebraic operations, and analyse abstract concepts. A breakdown in a single procedural element such as forgetting a derivative formula can often impede higher-order critical thinking (Baker, 2022). In Vygotskyan terms, AI functions as a More Knowledgeable Other (MKO) operating within the student's Zone of Proximal Development (ZPD) (Wang, 2024). By offloading routine information retrieval or initial drafting tasks to AI, students are able to liberate their working memory. This observation is consistent with recent findings on cognitive offloading (Gerlich, 2025b); rather than becoming bogged down by formulaic syntax or rigid definitions, students are empowered to focus on the underlying physical significance of the phenomena.

For example, instead of expending 15 minutes merely to recall a specific integral formula, a student can prompt the AI to retrieve it, subsequently dedicating that time to analyzing the integral's relevance to the specific physical boundary conditions. Here, GenAI serves effectively as an on-demand tutor or 'Socratic Partner.' As highlighted in international literature, AI tools like ChatGPT offer detailed, step by step elucidations of complex phenomena, such as force decomposition on an inclined plane (Liang et al., 2023). Furthermore, the immediate feedback provided by GenAI enables real-time rectification of misconceptions a distinct advantage over conventional classroom settings where lecturers are constrained in their ability to address individual student needs simultaneously (Borah et al., 2024).

3.3.2. Analysis of Critical Thinking Indicators

Critical thinking proficiency in physics confers significant advantages upon students. Specifically, the indicators of argumentation and deduction in physics demand the construction of robust logical chains (e.g., "If net force is zero, then acceleration is zero; consequently, velocity is constant"). As Large Language Models (LLMs) trained on extensive textual datasets, GenAI models excel at structuring logical arguments (Ciubotaru, 2025). Exposure to these high-quality logical structures even amidst occasional factual imperfections allows students to internalize sophisticated reasoning patterns, thereby bolstering their competencies in "Providing Arguments" and "Performing Deduction" (Indicators 2 & 3). The marked improvement observed in the experimental group underscores the potential of AI to cultivate evaluative judgment. When students task AI with solving a physics problem and are subsequently required to determine the solution's physical feasibility such as verifying units, limits, and vector directions they engage in high-level metacognition. This process corresponds directly to the "Evaluation" indicator (Indicator 5), a skill often challenging to in still via traditional lecturing. In this capacity, GenAI functions to mitigate procedural cognitive load.

Consistent with findings by Jarrahi et al (2023), AI operates as a collaborative partner, managing "information retrieval" and "initial structuring." Facilitated by structured Type A Student Worksheets, students in the experimental group could prompt AI to "elucidate the concept of normal force on an inclined plane" or "outline general solution steps." This delegation effectively liberates working memory capacity, enabling students to concentrate on higher-order cognitive activities: Evaluation and Synthesis. Moreover, these findings contest the pessimistic narrative that AI stifles creativity or critical acumen. The "AI Answer Validation" protocol implemented in this intervention emerged as a potent exercise in critical thinking. By requiring students to "verify whether ChatGPT's response aligns with Newton's Second Law in the textbook," the intervention compelled them to activate an analytical mindset. Transitioning from passive consumers to active evaluators, students engaged in a fault-finding process with the "smart machine," stimulating a sensitivity to physical nuances often missed in passive learning. This aligns with (Liang & Bai, 2025) who posit that dialectical interaction with AI deepens conceptual understanding through verification and reflection. Ultimately, the GenAI intervention appears to have successfully reduced extraneous cognitive load. By leveraging ChatGPT or DeepSeek to provide scaffolding for derivations or definitions, students were able to reallocate mental resources toward higher-order thinking: argument analysis, validity evaluation, and solution synthesis.

3.3.3. "The Double-Edged Sword": Balancing Cognitive Engagement and Dependency

These findings present a compelling counter-narrative to the "educational apocalypse" often predicted in the context of AI. Nevertheless, recent literature cautions that unchecked reliance on AI may precipitate a decline in independent problem-solving capabilities (Kobaba et al., 2025). While the statistical outcomes reveal a significant positive impact, it is crucial to observe that the experimental group's mean N-Gain plateaued below the "High" category, suggesting inherent limitations. It is plausible that a subset of students, despite achieving higher scores, succumbed to the temptation of cognitive "shortcuts" accepting AI responses without rigorous evaluation, particularly if their prompting behaviours were "answer-seeking" rather than "explanation-seeking". Furthermore, some experimental students may have fallen prey to the illusion of competence. The fluency of AI explanations can foster a false sense of mastery; yet, this understanding remains fragile. When confronted with tasks demanding independent problem-solving, the residue of this superficial understanding proves insufficient. If students fail to detect AI errors due to a deficit in foundational concepts, the resulting process is essentially the internalization of misconceptions. Consequently, the lecturer's role was not obsolete but rather pivoted to that of a facilitator and expert validator. The experimental group's success was largely attributed to the instructional design of the Student Worksheets (LKM), which mandated explicit cross-reference with textbooks to mitigate blind acceptance.

The Positive Edge (Benefits): Students in the AI group did not simply transcript answers; the PBL framework compelled them to evaluate AI outputs. The "Evaluation" component of critical thinking necessitates credibility assessment. By confronting students with AI-generated text known for its susceptibility to errors this intervention effectively converted the risk of AI "hallucinations" into a pedagogical asset. Students assumed the role of "editors" or "auditors," a capacity requiring a deeper grasp of the material than traditional composition (Kshetri, 2023). Mitigating the Negative Edge (Risks): The "perilous sharp edge" of this tool lies in potential dependency and the dulling of creativity. The instructional design served to blunt this risk. The worksheets required students to explicitly critique AI reasoning, thereby shifting the interaction from passive consumption to collaborative negotiation, aligning with contemporary human-AI collaboration frameworks. Without such structured guidance, critical thinking scores would likely have mirrored the lower outcomes observed in unstructured AI studies (Lawasi et al., 2024). Future inquiries should prioritize longitudinal studies to determine whether these gains in critical thinking are durable or if they lead to AI "addiction." Moreover, investigating the specific "prompt engineering" strategies employed by high-performing students could cultivate a repository of best practices for physics educators in Indonesia.

3.3.4. Implications for Science Education

These findings corroborate the learning trends articulated by Kasneci et al., (2023b) which highlight the transformative potential of Large Language Models (LLMs) like ChatGPT. By adapting content, pacing, and learning styles to individual student needs, AI functions as an ever-present personal tutor. Specifically, this study aligns with research indicating that structured AI interventions within PBL models bolster data analysis and problem-solving proficiency (Su, 2022). However, this study provides a nuanced contribution to the Indonesian higher education landscape. Amidst diverse student demographics and specific digital literacy hurdles, the successful implementation at UNDIKSHA proves that technological barriers are surmountable via robust pedagogical design. The observed enhancement in critical thinking signals that the quality gap in physics education can be bridged through prudent technology adoption. Moreover, the implications for national and institutional policy are unequivocal: a blanket ban on GenAI in academia represents a counter-productive step backward. Instead, universities are urged to embed "AI Literacy" into the core curriculum, instructing students in prompt engineering and, more importantly, the auditing and validation of algorithmic outputs essential skills for the 21st century.

Therefore, these findings necessitate significant practical shifts in the science education curriculum: First, the Evolution of the Lecturer: The traditional role of the lecturer as a sole information provider has been disrupted. Academics must evolve into "validation facilitators" and "logic coaches," guiding students through the deluge of AI-generated content and empowering them to discern valid physical truths from algorithmic hallucinations (Akhsan, 2023). Second, the Centrality of Prompt Engineering: The skill of crafting effective prompts must be circularized. Students need to understand that the caliber of the output is contingent upon the quality of the inquiry. A rudimentary prompt ("Solve this") produces instant answers that stifle cognition, while a sophisticated prompt ("Explain the underlying physics principles and provide an analogy") sparks conceptual understanding. Third, Process-Oriented Assessment: Evaluation paradigms must transition from assessing final outputs which AI can easily fabricate to scrutinizing the cognitive process, the rigor of argumentation, and critical evaluation capabilities.

3.3.5. Limitations of the Study

While the findings present compelling evidence for GenAI integration, several limitations must be acknowledged. First, the research was conducted with a relatively small sample size (N=38) and was confined to a specific institutional context within the Science Education program at Universitas Pendidikan Ganesha. This constraint may affect the broad generalizability of the findings to institutions with different student demographics or varying levels of digital infrastructure. Additionally, the intervention evaluated short term cognitive gains; it remains to be seen whether these improvements in critical thinking are durable over a longer academic period. Future studies should address these limitations by employing larger, multi-institutional cohorts and longitudinal designs to fully map the long-term cognitive impacts of AI dependency and literacy.

4. Conclusion

This research offers robust empirical evidence elucidating the dualistic nature of Generative AI in introductory Physics instruction: (1) Statistical Efficacy: The integration of GenAI within the Problem-Based Learning (PBL) framework demonstrated significant efficacy in enhancing critical thinking proficiency, yielding markedly superior outcomes (F=100.07; p<0.001) relative to conventional methods.; (2) The Double-Edged Sword Paradigm: GenAI operates as a double-edged sword. On the "sharp edge" (the beneficial side), it functions as intelligent scaffolding that facilitates problem decomposition and catalyzes active evaluation, propelling students toward the "Medium" improvement category. Conversely, on the "blunt edge" (the limiting side), the perils of AI hallucinations and cognitive dependency act as constraints that preclude the attainment of the "High" category. Without rigorous validation protocols, students remain susceptible to the illusion of competence.

Author Contributions

Putu Prima Juniartina contributed to the conceptualization, methodology, validation, and drafting of the introduction. Ni Luh Putu Mery Marlinda contributed to the introduction, results, and conclusion. I Made Oka Riawan was involved in the introduction, results, and discussion sections. Kadek Dwi Hendratma Gunawan contributed to the introduction, discussion, and reference compilation.

Funding

This research was funded by the DIPA Research Grant, Universitas Pendidikan Ganesha (Fiscal Year 2025).

Declaration of Conflicting Interests

This journal manuscript is the result of original research that has never been published before. The data used in this study comes from validly processed sources. The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Akhsan, J. (2023). Analysis of the teacher's role in educating students during the era of disruption. *Enigma in Education, 1*(1), 11–15. https://doi.org/10.61996/edu.v1i1.3

Alarbi, K., Halaweh, M., Tairab, H., Alsalhi, N. R., Annamalai, N., & Aldarmaki, F. (2024). Making a revolution in physics learning in high schools with ChatGPT: A case study in UAE. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(9), Article em2499. https://doi.org/10.29333/ejmste/14983

Alfarwan, A. (2025). Generative AI use in K–12 education: A systematic review. *Frontiers in Education, 10*, Article 1647573. https://doi.org/10.3389/feduc.2025.1647573

Ayu, H. D., Chusniyah, D. A., Kurniawati, M. P., Purwanti, P. F., Lukitawanti, S. D., & Putri, A. N. (2024). Problem-based learning (PBL) as an effective solution to enhance understanding of physics concepts: Systematic literature review. *Journal of Environment and Sustainability Education, 2*(2). https://doi.org/10.62672/joease.v2i2.29

Baker, B. (2022). Higher-level thinking in advanced physics courses. In *Critical thinking in the physics curriculum* (pp. 5-1–5-10). AIP Publishing. https://doi.org/10.1063/9780735424234_005

Bogdanova, D., & Snoeck, M. (2019). Use of personalized feedback reports in a blended conceptual modelling course. In *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)* (pp. 672–679). IEEE. https://doi.org/10.1109/MODELS-C.2019.00103

Borah, A. R., T. N., N., & Gupta, S. (2024). Improved learning based on GenAI. In *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)* (pp. 1527–1532). IEEE. https://doi.org/10.1109/IDCIoT59759.2024.10467943

Ciubotaru, B.-I. (2025). *Generative AI and large language models: A comprehensive scientific review*. *Preprints*. https://doi.org/10.20944/preprints202504.0413.v1

Ennis, R. H. (1985). *Developing minds: A resource book for teaching thinking* (A. L. Costa, Ed.). Association for Supervision and Curriculum Development.

Gerlich, M. (2025). From offloading to engagement: An experimental study on structured prompting and critical reasoning with generative AI. *Data, 10*(11), 172. https://doi.org/10.3390/data10110172

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Harahap, R., Sudarma, T. F., Novika, S., Mannasalwa, Z., & Putri, N. A. (2025). Effectiveness and innovation of problem-based learning in physics learning in a decade: A literature analysis of critical thinking development. *Jurnal Eduscience (JES), 12*(3), 813–825. https://doi.org/10.36987/jes.v12i3.7121

Helal, M. Y. I., Elgendy, I. A., Albashrawi, M. A., Dwivedi, Y. K., Al-Ahmadi, M. S., & Jeon, I. (2025). The impact of generative AI on critical thinking skills: A systematic review, conceptual framework and future research directions. *Information Discovery and Delivery*. Advance online publication. https://doi.org/10.1108/IDD-05-2025-0125

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons, 66*(1), 87–99. https://doi.org/10.1016/j.bushor.2022.03.002

Jho, H. (2024). Leveraging generative AI in physics education: Addressing hallucination issues in large language models. *New Physics: Sae Mulli, 74*(8), 812–823. https://doi.org/10.3938/NPSM.74.812

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Kasneci, G., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khan, N., & Saunderson, S. (2024). Generative artificial intelligence in education: A narrative literature review. *World Journal on Educational Technology: Current Issues, 16*(2), 112–126. https://doi.org/10.18844/wjet.v16i2.9217

Kobaba, B. M., Lihega, J. K., & Khalfan, K. A. (2025). AI in higher education: Effect on student competency and independent learning. *Asian Journal of Education and Social Studies, 51*(7), 1100–1106. https://doi.org/10.9734/AJESS/2025/v51i72194

Kotsis, K. T. (2025). ChatGPT and DeepSeek evaluate one another for science education. *EIKI Journal of Effective Teaching Methods, 3*(1). https://doi.org/10.59652/jetm.v3i1.439

Kshetri, N. (2023). The future of education: Generative artificial intelligence's collaborative role with teachers. *IT Professional, 25*(6), 8–12. https://doi.org/10.1109/MITP.2023.3333070

Lawasi, M. C., Rohman, V. A., & Shoreamanis, M. (2024). The use of AI in improving students' critical thinking skills. *Proceedings Series on Social Sciences & Humanities, 18*, 366–370. https://doi.org/10.30595/pssh.v18i.1279

Liang, E. S., & Bai, S. (2025). Generative AI and the future of connectivist learning in higher education. *Journal of Asian Public Policy, 18*(2), 329–351. https://doi.org/10.1080/17516234.2024.2386085

Liang, Y., Zou, D., Xie, H., & Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learning Environments, 10*(1), Article 52. https://doi.org/10.1186/s40561-023-00273-7

Mahligawati, F., Allanas, E., Butarbutar, M. H., & Nordin, N. A. N. (2023). Artificial intelligence in physics education: A comprehensive literature review. *Journal of Physics: Conference Series, 2596*(1), 012080. https://doi.org/10.1088/1742-6596/2596/1/012080

Nemalynne, A. A., Thelma, C. D., Shermaine, A. D., Jessica, R. L., Romelyn, T. L., Ryan, B. T., Nove Lheen, C. T., & Eisle Keith, R. T. (2023). Challenges in teaching science and its transition to post-pandemic education. *American Journal of Multidisciplinary Research and Innovation, 2*(3), 15–22. https://doi.org/10.54536/ajmri.v2i3.1195

Şengül, Ö. (2023). Reconsidering physics education: Barriers to student learning and ways to effective physics teaching. In *Eğitim Bilimleri Araştırmaları V*. Özgür Yayınları. https://doi.org/10.58830/ozgur.pub171.c855

Su, K.-D. (2022). Implementation of innovative artificial intelligence cognitions with problem-based learning guided tasks to enhance students' performance in science. *Journal of Baltic Science Education, 21*(2), 245–257. https://doi.org/10.33225/jbse/22.21.245

Wang, Y. (2024). Integration of AI technology and ZPD: Pinpointing the zone of student development to promote educational innovation and efficiency improvement. *Journal of Education, Humanities and Social Sciences, 38*, 99–104. https://doi.org/10.54097/wws14m21

Worthington, P. (2019). Development and validation of a critical thinking framework. *SURG Journal, 11*. https://doi.org/10.21083/surg.v11i0.5189

Wu, X., & Valente, D. (2020). Adapting differentiated cognitive load measurement in physics classroom. In *Physics Education Research Conference Proceedings 2019* (pp. 663–668). American Association of Physics Teachers. https://doi.org/10.1119/perc.2019.pr.Wu

Younis, A. (2025). Advancing learning outcomes in physics education through artificial intelligence integration. *European Journal of Education and Pedagogy, 6*(5), 34–39. https://doi.org/10.24018/ejedu.2025.6.5.996