



Evaluating Artificial Intelligence in Solving Physics Problems: A Comprehensive Analysis of Essay and Multiple-Choice Questions on Electrical Circuits

Aji Saputra^{1*}, Muliadi², Hutri Handayani Isra¹

¹ Physics Education, Faculty of Teacher Training and Education, Universitas Khairun, Ternate, Indonesia.

² Chemistry Education, Faculty of Teacher Training and Education, Universitas Khairun, Ternate, Indonesia.

Received: January 02, 2026

Revised: February 10, 2026

Accepted: March 25, 2026

Published: March 31, 2026

Corresponding Author:

Aji Saputra

aji.saputra@unkhair.ac.id

DOI: [10.29303/jppipa.v12i3.14494](https://doi.org/10.29303/jppipa.v12i3.14494)

 Open Access

© 2026 The Authors. This article is distributed under a (CC-BY License)



Abstract: This study aims to examine the performance of artificial intelligence, specifically ChatGPT, in solving physics problems on electrical circuits by comparing essay and multiple-choice formats. The study employed a research and development approach combined with descriptive qualitative analysis to develop and validate equivalent test items with identical content and cognitive demands. The validated items were administered to ChatGPT, and its responses were analyzed based on accuracy and problem-solving processes. The results show that ChatGPT achieved higher accuracy in multiple-choice questions (100%) compared to essay questions (60%). Errors in essay responses were primarily conceptual and occurred in image-based and structurally complex circuit problems, particularly during the initial interpretation stage. These findings indicate that question format significantly influences AI performance, with multiple-choice questions promoting more structured reasoning while potentially overestimating AI capability. In conclusion, combining essay and multiple-choice formats provides a more comprehensive evaluation of AI problem-solving performance in physics education.

Keywords: Artificial intelligence; ChatGPT; Electrical circuits; Essay question; Multiple-choice question

Introduction

The development of Artificial Intelligence (AI) technology in recent years has progressed rapidly (Niu et al., 2025) and has influenced various sectors of life, including education (Bates et al., 2020; Crompton & Burke, 2023; Wang et al., 2024). In educational contexts, AI has been widely utilized to support learning processes for students, teachers, and other educational practitioners (Agyare et al., 2025). AI enables users to access information, engage in interactive discussions, and solve complex learning problems (Berliana et al., 2024). One of the most widely adopted AI applications in education is AI-based chatbots, such as ChatGPT (Shahzad et al., 2024), which allow students to obtain answers to various questions, including physics problems that require mathematical calculations and

deep conceptual understanding (Agyare et al., 2025). Furthermore, AI can be implemented in intelligent tutoring systems (ITSs) and demonstrates strong potential to support sustainable education (Lin et al., 2023).

The rapid advancement of AI has also reshaped research in science education. Bibliometric studies indicate a substantial increase in AI-related publications, highlighting the growing importance of generative AI in teaching, learning, and assessment processes (Ayuni et al., 2024). Numerous studies have highlighted the potential of AI as a supportive tool in educational practice. A systematic review by Prananta et al. (2023) showed that ChatGPT is predominantly integrated into science learning media to support information retrieval and scientific communication. Empirical studies further demonstrated that AI-assisted learning environments

How to Cite:

Saputra, A., Muliadi, & Isra, H. H. (2026). Evaluating Artificial Intelligence in Solving Physics Problems: A Comprehensive Analysis of Essay and Multiple-Choice Questions on Electrical Circuits. *Jurnal Penelitian Pendidikan IPA*, 12(3), 667-689. <https://doi.org/10.29303/jppipa.v12i3.14494>

enhance student engagement and learning outcomes (Hariadi et al., 2025). In terms of instructional materials, AI-based e-learning modules and AI-assisted e-modules have been reported to improve students' critical thinking skills, communication abilities, and conceptual understanding in science and engineering contexts (Aji et al., 2025; Pairunan et al., 2025; Purba et al., 2025). Overall, these studies position AI primarily as a facilitator of learning when embedded within instructional design.

Beyond its role in learning facilitation, AI has been examined across various educational domains. Research in higher education and teacher education has explored AI adoption, AI literacy among educators, and its role in improving access, learning outcomes, and cost efficiency, although acceptance remains limited and influenced by technological trust and perceived usefulness (Sperling et al., 2024). Studies on learning and assessment indicate that generative AI can support formal academic writing through immediate feedback but is less effective for informal tasks, while also presenting opportunities to address academic integrity challenges (Azap, 2025; Özçelik & Ekşi, 2024; Xia et al., 2024). In adaptive and personalized learning, AI applications such as intelligent tutoring systems provide personalized learning experiences and real-time feedback, improving learning outcomes (Wang et al., 2024; Zawacki-Richter et al., 2024). AI has also been applied in educational management and STEM education to support learning analytics and adaptive instruction aligned with labor market needs (Prasetya et al., 2025; Salas-Pilco & Yang, 2022; Smith-Mutegi et al., 2025). Despite these benefits, ethical concerns related to trust, misuse, creativity, and critical thinking remain significant challenges, highlighting the importance of educator involvement in guiding AI use (Nam & Bai, 2023; Schmidt et al., 2025).

Empirical evidence suggests that the accuracy of AI-generated answers varies depending on the complexity of the physics problems. Nurhuda et al. (2023) reported that ChatGPT accurately solved image-based physics problems involving resistor circuits, particularly when the problems involved relatively simple circuit configurations. However, Firman et al. (2023) found that AI responses tend to be less accurate when addressing physics problems that require deeper conceptual understanding and more complex application of formulas. These findings indicate that while AI can perform well on simple, structured physics problems, its effectiveness is constrained when faced with tasks that require complex reasoning and in-depth conceptual analysis. In addition, recent studies have explored AI as a tool for diagnosing student misconceptions, positioning it as an evaluator rather than a subject of evaluation (Purnama et al., 2025).

Despite the extensive application of AI in education, systematic investigations that rigorously examine the conceptual accuracy and scientific validity of AI-generated answers remain limited, especially in the context of physics education. Furthermore, existing studies often lack validated assessment instruments and rarely compare different question formats, such as essay and multiple-choice, which may influence AI performance.

From a theoretical perspective, problem solving in physics requires not only procedural knowledge but also deep conceptual understanding and reasoning skills. Furthermore, assessment theory suggests that different question formats, such as essay and multiple-choice, measure different levels of cognitive processes and reasoning abilities (Zulpan et al., 2024). Therefore, this study aims to address this gap by systematically analyzing the performance of AI, specifically ChatGPT, in answering physics questions involving electrical circuits. This research compares AI performance across essay and multiple-choice question formats using a validated and reliable assessment instrument. The instrument is developed through rigorous testing procedures, including analyses of content validity, item validity, reliability, the discrimination index, and difficulty level. By examining not only the accuracy of AI-generated answers but also the reasoning processes underlying them, this study seeks to deepen understanding of AI's potential and limitations in solving complex physics problems and to contribute to the development of more effective technology-based assessment tools in physics education.

Method

Time and Location of the Research

This study was conducted in 2025 at SMA Negeri 4 Ternate, Indonesia. Thirty students participated as respondents for the purpose of item validity and reliability testing. The analysis of problem-solving ability was carried out using ChatGPT, which served as the primary subject of evaluation.

Research Methods

This study employed a Research and Development (R&D) approach combined with a descriptive qualitative method. The R&D approach was selected to develop a valid and reliable test instrument (Sugiyono, 2021), while the descriptive qualitative method was used to analyze the ability of artificial intelligence to solve physics problems related to electrical circuit topics. The study consisted of several stages, each designed to ensure the validity of the test instrument and to evaluate the effectiveness of AI in solving the developed physics problems.

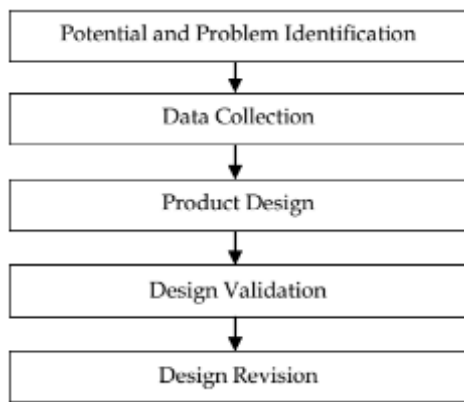


Figure 1. Flowchart of R&D procedures

The R&D procedures implemented in this study are illustrated in Figure 1, consisting of five stages: (1) potential and problem identification, identifying key issues and needs; (2) data collection, gathering relevant information; (3) product design, developing the initial prototype; (4) design validation, evaluating feasibility and appropriateness; and (5) design revision, refining the product based on feedback. This limitation was applied because the primary objective of the study was to ensure that the test items were valid prior to being administered to ChatGPT. According to Sugiyono (2021), R&D research conducted up to the design validation stage focuses on product development or refinement, with the main objective of ensuring that the product design meets predetermined criteria and requirements before further testing.

Research Stages

The research was conducted through several systematic stages. First, test items on electrical circuits were developed, covering concepts such as series circuits, parallel circuits, Ohm’s law, and the Wheatstone bridge. The items were designed in two formats, namely essay and multiple-choice, with identical content and cognitive demands. Second, content validity testing was conducted by six experts using the Content Validity Index (CVI) method to evaluate the relevance of the items in terms of construction, scientific content, and language. (Puspitasari & Febrinita, 2021). Third, empirical testing was conducted involving 30 students to analyze item validity, reliability, discrimination index, and difficulty level using SPSS software. Finally, validated test items were administered to ChatGPT. The AI was asked to respond to both essay and multiple-choice questions without additional prompts, and its responses were recorded for further analysis.

Data Analysis

The CVI method was used to assess the extent to which the test items were relevant and valid in terms of construction, scientific content, and language clarity (Zulpan et al., 2024). In the content validity assessment, 12 evaluation statements were rated by six experts across three aspects: construction, scientific content, and language. The indicators are presented in Table 1.

Table 1. Content Validity Evaluation Statements

Aspect	Statements
Construction	The instructions for answering the questions are clear
	The questions are arranged logically and systematically
	The questions do not lead students toward a particular correct answer
	The questions are independent and do not depend on one another
Scientific Content	The questions use appropriate physics terminology and symbols correctly
	The questions assess students’ basic understanding of Ohm’s law
	The questions assess students’ understanding of series circuits
	The questions assess students’ understanding of parallel circuits
	The questions assess students’ understanding of combined (mixed) circuits
Language	The questions assess students’ understanding of the Wheatstone bridge
	The language used in the questions is clear and unambiguous
	The questions do not contain elements related to ethnicity, religion, race, or intergroup discrimination

The measurement scale used in the expert validation sheet was ordinal, ranging from 1 to 4. The selection of a four-point scale was intended to avoid neutral responses that may occur at the midpoint. The descriptors applied to this scale were as follows: 1 = not relevant, 2 = less relevant, 3 = moderately relevant, and 4 = highly relevant. The ordinal scale was subsequently converted to a dichotomous 0/1 scale for analysis using the Content Validity Index (CVI) approach. Ratings of 1 and 2 were classified as not acceptable (0), whereas

ratings of 3 and 4 were classified as acceptable (1) (Yusof, 2019).

The experts’ ratings were then aggregated. Instrument validity testing using the CVI approach was conducted by calculating two indices: the item-level content validity index (I-CVI) and the scale-level content validity index (S-CVI). The determination of content validity was based on the criteria presented in Table 2.

Table 2. Criteria for Determining Content Validity Using the CVI Approach (Puspitasari & Febrinita, 2021)

Number of Experts	Acceptable CVI Value
2 experts	Minimum 0.80
3 - 5 experts	Must be 1.00
≥ 6 experts	Minimum 0.83
6 - 8 experts	Minimum 0.83
≥ 9 experts	Minimum 0.78

Subsequently, analyses of item validity, reliability, discrimination index, and difficulty level were conducted for the multiple-choice questions. For the essay questions, only item validity and reliability analyses were performed. The sample consisted of 30 students from SMA Negeri 4 Ternate City. Item validity testing was conducted to ensure that each test item was highly relevant to the assessed topic, thereby guaranteeing the quality of the questions.

To ensure the instrument’s consistency and accuracy, internal consistency was assessed using an internal consistency measure to confirm that the test items produced consistent results across different groups of respondents. Furthermore, analyses of the discrimination index and difficulty level were conducted using a small group of respondents to assess whether the test items had sufficient variation to distinguish among students with different ability levels. These analyses ensured that the developed items covered appropriate levels of difficulty in accordance with the expected competencies (Sugiyono, 2021). All four analyses, namely item validity, reliability, discrimination index, and difficulty level, were performed using SPSS software.

Validity testing was conducted using the following decision criteria: if the Sig. (2-tailed) If the value was less than 0.05, the item was classified as valid; otherwise, it was classified as invalid. (2-tailed) value was greater than or equal to 0.05, the item was classified as not valid. Reliability testing was performed using Cronbach’s Alpha, with the decision criteria that the instrument was considered reliable if the calculated Cronbach’s Alpha value was greater than or equal to the reference Cronbach’s Alpha value, and not reliable if the calculated value was less than the reference value. The classification of item reliability categories is presented in Table 3.

Table 3. Categories of Item Reliability (Cahyaningrum et.al., 2023)

Coefficient Value	Category
0.000 - 0.199	Very Low
0.200 - 0.399	Low
0.400 - 0.599	Moderate
0.600 - 0.799	High
0.800 - 1.000	Very High

The discrimination index for the test items was categorized based on the calculated r value (as shown in SPSS output), as presented in Table 4.

Table 4. Categories of Item Discrimination Index (Mutakin, 2023)

Coefficient Value	Category
0.40 - 1.00	Good
0.30 - 0.39	Moderate (no revision required)
0.20 - 0.29	Fair (revision required)
-1.00 - 0.19	Poor

The difficulty level of the test items was categorized based on the mean score (as shown in Table 5).

Table 5. Categories of Item Difficulty Level (Mutakin, 2023)

Mean	Category
0.00 - 0.25	Difficult
0.26 - 0.75	Moderate
0.76 - 1.00	Easy

After developing a valid and reliable test instrument, data were collected using a structured procedure. The validated test items were administered directly to an artificial intelligence system, specifically ChatGPT, without any additional or customized prompts. The questions were copied verbatim, allowing ChatGPT to respond solely based on the instructions embedded in each test item. ChatGPT was asked to answer both multiple-choice and essay questions on electrical circuits, and each item was administered once.

The AI-generated responses were then analyzed using a descriptive qualitative approach. The accuracy of each response was evaluated using a dichotomous scoring scheme, in which answers were classified as either correct or incorrect based on established and authoritative physics literature. In addition to accuracy, the problem-solving processes presented by ChatGPT were examined to assess whether the AI followed appropriate procedures and provided clear, logical explanations, particularly for items requiring more complex reasoning. Furthermore, items that were answered incorrectly were analyzed to identify patterns in AI limitations, particularly those related to conceptual understanding and problem complexity. The findings of this analysis were used to formulate recommendations to improve item design and assessment strategies for evaluating AI performance in physics education.

Result and Discussion

Development and Content Validity of the Instrument

The development of the test instrument in this study began with 15 essay and multiple-choice items

designed to assess the same content and concepts, differing only in format. This approach minimized bias by ensuring that any differences in outcomes were due to the question format rather than variations in content or cognitive demands.

The next stage involved content validity testing of the developed items. The results of the content validity evaluation conducted by six experts (V1-V6) are presented in Table 6 for the essay items and in Table 7 for the multiple-choice items.

Table 6. Results of Content Validity Assessment for Essay Items

Aspect	Relevance Proportion						Mean i-CVI
	V1	V2	V3	V4	V5	V6	
Construction	1.00	0.75	1.00	0.75	1.00	1.00	0.92
Scientific Content	1.00	1.00	1.00	0.75	1.00	0.75	0.92
Language	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	s-CVI						0.94

Table 7. Results of Content Validity Assessment for Multiple-Choice Items

Aspect	Relevance Proportion						Mean i-CVI
	V1	V2	V3	V4	V5	V6	
Construction	1.00	0.75	1.00	0.75	1.00	1.00	0.92
Scientific Content	1.00	1.00	1.00	0.75	1.00	0.75	0.92
Language	1.00	1.00	1.00	1.00	1.00	0.75	0.96
	s-CVI						0.93

The content validity analysis yielded an S-CVI value of 0.94 for the essay items and 0.93 for the multiple-choice items. Based on the content validity criteria presented in Table 2, both sets of items are valid and acceptable for use in this study.

Item Validity and Reliability Analysis

The subsequent stage involved item validity testing. This analysis was conducted to ensure that each test item accurately measured the intended construct in accordance with the predefined competency indicators. The results of the item validity analysis are presented in Table 8 for the essay items and in Table 9 for the multiple-choice items.

The results presented in Tables 8 and 9 indicate the same conclusion: items numbered 1, 6, 7, 8, and 11 were classified as not valid in both the essay and multiple-choice formats. This finding is consistent with the test instrument’s design: the content of the questions was identical, differing only in format, with one set presented as essay items and the other as multiple-choice items.

Subsequently, the valid items were subjected to reliability testing using SPSS through the Analyze – Scale – Reliability Analysis procedure. The Cronbach’s Alpha coefficients obtained from this analysis are presented in Table 10. Table 10 shows that the reliability

level of both the essay and multiple-choice items was classified as very high.

Table 8. Results of Item Validity Analysis for Essay Items

Item	Sig.(2-tailed)	Remarks	Pearson	Interpretation
1	0.248	Not valid	0.218	-
2	0.001	Valid	0.627	High
3	0.004	Valid	0.507	Moderate
4	0.002	Valid	0.534	Moderate
5	0.001	Valid	0.654	High
6	0.356	Not valid	0.175	-
7	0.064	Not valid	0.342	-
8	0.850	Not valid	0.036	-
9	0.001	Valid	0.569	Moderate
10	0.001	Valid	0.653	High
11	0.123	Not valid	0.288	-
12	0.001	Valid	0.718	High
13	0.001	Valid	0.571	Moderate
14	0.001	Valid	0.718	High
15	0.001	Valid	0.569	Moderate

Table 9. Results of Item Validity Analysis for Multiple-Choice Items

Item	Sig.(2-tailed)	Remarks	Pearson	Interpretation
1	0.199	Not valid	0.241	-
2	0.001	Valid	0.688	High
3	0.007	Valid	0.485	Moderate
4	0.010	Valid	0.463	Moderate
5	0.001	Valid	0.679	High
6	0.223	Not valid	0.229	-
7	0.859	Not valid	-0.034	-
8	0.199	Not valid	0.241	-
9	0.004	Valid	0.510	Moderate
10	0.001	Valid	0.615	High
11	0.716	Not valid	0.069	-
12	0.001	Valid	0.618	High
13	0.004	Valid	0.514	Moderate
14	0.001	Valid	0.799	High
15	0.012	Valid	0.453	Moderate

Table 10. Results of Reliability Testing

Item Type	Cronbach’s Alpha
Essay	0.811
Multiple-Choice	0.804

Discrimination Index and Difficulty Level

The discrimination index and difficulty level analyses were conducted only for the multiple-choice items. Essay items produce continuous scores based on scoring rubrics, unlike the dichotomous nature of multiple-choice items. Therefore, the evaluation of essay item quality focused primarily on validity and reliability as the main indicators of measurement quality, whereas analyses of classical discrimination index and difficulty level based on the proportion of correct responses were not applied. The results of the item discrimination index analysis are presented in Table 11.

Table 11. Results of Item Discrimination Index Analysis

Valid Items	R	Discrimination Index
2	0.079	Poor
3	0.000	Poor
4	-0.177	Poor
5	0.250	(Revision Required)
9	0.101	Poor
10	0.177	Poor
12	0.301	Moderate
13	0.302	Moderate
14	0.269	(Revision Required)
15	0.999	Good

The results of the discrimination index analysis indicate that several test items exhibited poor discrimination. This condition was influenced by the limited number of respondents (N = 30) and the characteristics of the test items, which were designed to evaluate AI performance rather than to select or rank students. Therefore, the discrimination index analysis was used descriptively to inform item selection, rather than as the sole criterion for item elimination. The difficulty level of the items is presented in Table 12.

Table 12. Results of Item Difficulty Level Analysis

Valid Items	Mean	Difficulty Level
2	0.767	Easy
3	0.833	Easy
4	0.667	Moderate
5	0.800	Easy
9	0.567	Moderate
10	0.667	Moderate
12	0.553	Moderate
13	0.733	Moderate
14	0.567	Moderate
15	0.200	Difficult

Table 12 shows that the test items ranged from easy to difficult. Most items were classified as moderate in difficulty, specifically items numbered 4, 9, 10, 12, 13, and 14. Items categorized as easy were 2, 3, and 5, whereas item 15 was classified as difficult and had the lowest mean score.

After obtaining a set of test items that met the validity and reliability criteria, the subsequent analysis focused on examining the AI-generated responses to these selected items. All validated items were administered to the AI for analysis. The results are first presented for the essay questions, as this format allows for a more detailed examination of the AI's reasoning processes and the quality of its explanations.

AI Performance on Essay Questions

Question 1 (Essay)

Three resistors are connected in parallel (R1=15 Ω; R2=10 Ω; R3=6 Ω). The resistors are connected to a voltage source of 9.3 V. Determine the electric current

and the voltage across R1, respectively, expressed in amperes and volts.

ChatGPT's Response

Step 1: Calculating the equivalent resistance (R_e) of the parallel circuit.

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

$$R_{eq} = 3\Omega \tag{1}$$

Step 2: Calculating the total current

$$I_{total} = \frac{V}{R_{eq}}$$

$$I_1 = 3.1 A \tag{2}$$

Step 3: Calculating the current through R₁

Since the circuit is parallel, the voltage across each resistor is the same as the source voltage. Therefore,

$$I_1 = \frac{V}{R_1}$$

$$I_1 = 0.62 A \tag{3}$$

Thus, the current through resistor R₁ is 0.62 A, and the voltage across R₁ is 9.3 V.

For Question 1, ChatGPT's solution is correct and consistent with established physics principles. However, the response includes an unnecessary procedural step, specifically the calculation of the total current in Step 2. The question explicitly asked only for the current and voltage across R₁. According to parallel circuit theory, the voltage across each branch is equal to the source voltage, and the current through R₁ can be directly determined using Ohm's law, as shown in Step 3. Therefore, the inclusion of Step 2 represents a procedural redundancy rather than a conceptual error.

Question 2 (Essay)

Two resistors, R1=6 Ω and R2=3 Ω, are connected in parallel. The parallel combination is then connected in series with a third resistor, R3=8 Ω. When the circuit is connected to a 12 V voltage source, determine the electric current through and the voltage across R1, respectively, expressed in amperes and volts.

ChatGPT's Response

Step 1: Calculating the equivalent resistance (R_e) of resistors R₁ and R₂ connected in parallel

$$\frac{1}{R_{parallel}} = \frac{1}{R_1} + \frac{1}{R_2}$$

$$R_{parallel} = 2\Omega \tag{4}$$

Step 2: Adding resistor R₃ connected in series with the parallel combination of R₁ and R₂

$$R_{total} = 2\Omega + 8\Omega$$

$$R_{total} = 10\Omega \tag{5}$$

Step 3: Calculating the voltage and current through resistor R₁

Since resistors R₁ and R₂ are connected in parallel, the voltage across R₁ is equal to the voltage across the parallel combination, which can be determined using voltage division in the series circuit.

Summary of results for this problem:

- The equivalent resistance ($R_{parallel}$) of resistors R_1 and R_2 connected in parallel is $2\ \Omega$.
- The total resistance (R_{total}) of the circuit is $10\ \Omega$.
- The total current flowing in the circuit is $1.2\ A$.
- The voltage across the parallel combination (R_1 and R_2) is $2.4\ V$.
- The current flowing through resistor R_1 is approximately $0.4\ A$.

Although the final answers provided by ChatGPT for Question 2 are correct and consistent with established physics principles, the solution lacks procedural completeness. Specifically, the intermediate calculations used to determine the voltage across the parallel combination and the current through resistor R_1 were not presented explicitly. This indicates that while ChatGPT can produce correct numerical results, it does not always provide a transparent, fully articulated problem-solving process, which may limit the usefulness of its responses as a learning resource, particularly for students who rely on step-by-step explanations.

Question 3 (Essay)

The third question was presented in the form of an image; therefore, the researcher uploaded the question to ChatGPT as an image.

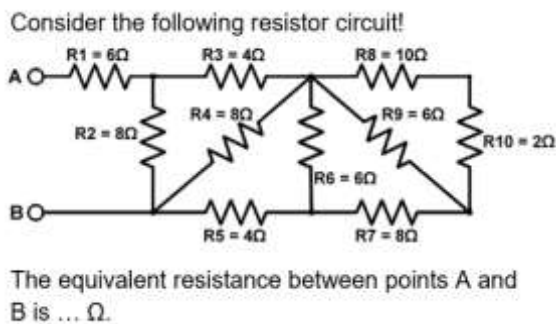


Figure 2. Question number 3

ChatGPT's Response

To find the equivalent resistance of the circuit between points A and B, let's break the circuit into smaller parts and analyze the resistors step by step:

Step 1: Identify Series and Parallel Connections

- Resistors R_1 and R_2 are in series
 $R_{12} = R_1 + R_2 = 14\ \Omega$ (5)
- Resistors R_3 and R_4 are in parallel
 $R_{34} = \frac{R_3 \cdot R_4}{R_3 + R_4} = 2.67\ \Omega$ (6)
- Resistors R_5 and R_6 are in series
 $R_{56} = R_5 + R_6 = 10\ \Omega$ (7)
- Resistors R_7 and R_8 are in parallel
 $R_{78} = \frac{R_7 \cdot R_8}{R_7 + R_8} = 4.44\ \Omega$ (8)

Step 2: Combine Sub-Circuits

- Combine R_{34} and R_{78} in series with R_9
 $R_{34,78,9} = R_{34} + R_{78} + R_9 = 13.11\ \Omega$ (9)

- Combine R_{12} and R_{56} in parallel with $R_{34,78,9}$
 $R_{total} = \frac{R_{12} \cdot R_{56}}{R_{12} + R_{56}} = 5.83\ \Omega$ (10)

Step 3: Add Remaining Resistor R_{10}

- Finally, add R_{10} in series
 $R_{final} = R_{total} + R_{10} = 7.83\ \Omega$ (11)

Therefore, the equivalent resistance between points A and B is approximately $7.83\ \Omega$.

For Question 3, ChatGPT's error can be classified as a conceptual error, as it incorrectly identified the series and parallel relationships between resistors at the initial stage of the analysis. The misclassification of resistors R_1 and R_2 as being connected in series, despite the presence of a branching path involving R_3 , indicates a fundamental misunderstanding of the circuit topology. This conceptual error occurred before any numerical calculation was performed, leading to an incorrect overall solution.

Notably, this question was presented as an image, suggesting that visual input may increase the likelihood of misinterpretation by AI, particularly in complex circuit configurations. Unlike Questions 1 and 2, which showed correct conceptual understanding but exhibited procedural redundancy or incompleteness, Question 3 reveals limitations in the AI's ability to accurately interpret and analyze graphical representations of electrical circuits. This finding highlights that AI performance in physics problem solving is influenced not only by mathematical complexity but also by input modality and problem structure.

The circuit configuration in Question 3 is more complex than those in Questions 1 and 2. Upon careful inspection of the circuit, the calculation can be simplified by starting with resistors R_8 and R_{10} , which are connected in series.

- $R_{8,10} = R_8 + R_{10} = 12\ \Omega$ (12)

- Resistors $R_{8,10}$ and R_9 are in parallel
 $R_{8,10,9} = \frac{R_{8,10} \cdot R_9}{R_{8,10} + R_9} = 4\ \Omega$ (13)

- Resistors $R_{8,10,9}$ and R_7 are in series
 $R_{8,10,9,7} = R_{8,10,9} + R_7 = 12\ \Omega$ (14)

- Resistors $R_{8,10,9,7}$ and R_6 are in parallel
 $R_{8,10,9,7,6} = \frac{R_{8,10,9,7} \cdot R_6}{R_{8,10,9,7} + R_6} = 4\ \Omega$ (15)

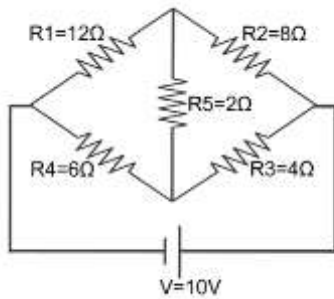
The subsequent steps follow the same procedures described in Equations (12)–(15).

- Resistors $R_{8,10,9,7,6}$ and R_5 are in series.
- Resistors $R_{8,10,9,7,6,5}$ and R_4 are in parallel.
- Resistors $R_{8,10,9,7,6,5,4}$ and R_3 are in series.
- Resistors $R_{8,10,9,7,6,5,4,3}$ and R_2 are in parallel
- Finally, resistors $R_{8,10,9,7,6,5,4,3,2}$ and R_1 are connected in series, yielding an equivalent resistance of $10\ \Omega$.

Question 4 (Essay)

The fourth question was presented in the form of an image; therefore, the researcher uploaded the question to ChatGPT as an image.

Consider the following resistor circuit!



The current flowing through R5 is ... A.

Figure 3. Question number 2

ChatGPT's Response

To solve for the current flowing through, we need to first determine the equivalent resistance of the entire circuit.

Step 1: Identify the resistors in series and parallel

- R1 = 12 Ω
- R2 = 8 Ω
- R3 = 4 Ω
- R4 = 6 Ω
- R5 = 2 Ω

Step 2: Simplify the circuit

Resistors R2 and R3 are in parallel

$$R_{2,3} = \frac{R_2 \cdot R_3}{R_2 + R_3} = 2.67\Omega \tag{16}$$

Combine R_{2,3} with R5 (in series)

$$R_{eq1} = R_{2,3} + R_5 = 4.67\Omega \tag{17}$$

For Question 4, the error committed by ChatGPT can be categorized as a conceptual pattern-recognition error, specifically related to its inability to identify the characteristic structure of a Wheatstone bridge within the resistor network. The incorrect assumption that resistors R₂ and R₃ could be combined directly in parallel and then connected in series with R₅ indicates a failure to recognize the overall circuit topology. As a result, the circuit simplification process was fundamentally flawed from the outset, leading to an incorrect equivalent resistance and an invalid final solution.

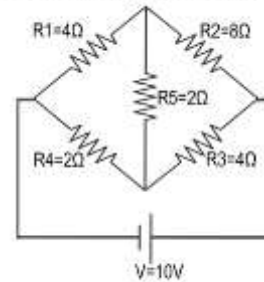
Similar to the findings in Question 3, this error occurred in a problem presented graphically, further suggesting that image-based inputs increase the likelihood of conceptual misinterpretation by AI, particularly in circuits with complex or nontrivial structures. In contrast to textual problems, which allow ChatGPT to rely primarily on symbolic manipulation, graphical circuit problems require accurate visual parsing and structural recognition. This limitation highlights that current AI systems may struggle to

recognize standard circuit patterns, such as Wheatstone bridges, thereby reducing their reliability in solving visually complex physics problems.

Question 5 (Essay)

The fifth question was presented in the form of an image; therefore, the researcher uploaded the question to ChatGPT as an image.

Observe the Wheatstone bridge circuit below!



The current flowing through R5 is ... A.

Figure 4. Question number 5

ChatGPT's Response

This is a Wheatstone bridge circuit. To find the current flowing through R5, we need to analyze the circuit step by step.

Check if the bridge is balanced: A Wheatstone bridge is balanced if:

$$\frac{R_1}{R_2} = \frac{R_4}{R_3} \tag{18}$$

Let's check this:

$$\frac{R_1}{R_2} = 0.5 \text{ and } \frac{R_4}{R_3} = 0.5 \tag{19}$$

Since the ratios are equal, the bridge is balanced.

Determine the current through R5: In a balanced Wheatstone bridge, no current flows through R5. Thus, the current through R5 is 0 A.

For Question 5, ChatGPT's solution is correct and demonstrates an appropriate understanding of the Wheatstone bridge concept. ChatGPT successfully identified the circuit as a Wheatstone bridge, unlike in Question 4, where it failed to recognize the Wheatstone bridge configuration.

A more detailed comparison reveals that the circuit structures in Questions 4 and 5 are highly similar, differing only in the resistor values. However, the responses generated by ChatGPT differ substantially. An examination of the question prompts shows that Question 5 explicitly included the instruction "Observe the Wheatstone bridge circuit below!" This explicit cue informed ChatGPT that the circuit presented was a Wheatstone bridge.

These findings indicate that explicit instructions or prompts strongly influence AI performance, particularly in guiding ChatGPT to correctly recognize circuit patterns. This result suggests that ChatGPT's ability to

solve physics problems depends not only on the circuit's structural complexity but also on the clarity and specificity of the instructions in the problem statement.

The analysis subsequently focused on the multiple-choice questions. The multiple-choice items analyzed addressed the same problems as the essay questions but were presented in a multiple-choice format. This approach was adopted because the items that met the validity criteria in the previous analysis were identical in content, allowing for an equitable comparison between question formats. Consequently, the only distinguishing factor among the analyzed items was the question format.

AI Performance on Multiple-Choice Questions

Question 1 (Multiple-Choice)

Three resistors are connected in parallel ($R_1=15 \Omega$, $R_2=10 \Omega$, $R_3=6 \Omega$). The resistors are connected to a voltage source of 9.3 V. The electric current through and the voltage across R_1 are, respectively:

- A. 9.3 A and 9.3 V
- B. 6.2 A and 6.2 V
- C. 5 A and 5 V
- D. 3.1 A and 6.2 V
- E. 0.62 A and 9.3 V

ChatGPT's Response

Principle of a parallel circuit

The voltage across each resistor is equal to the source voltage:

$$V_{R1}=9.3 \text{ V} \tag{20}$$

The current through each resistor is determined using Ohm's law:

$$I_1 = \frac{V}{R_1}$$

$$I_1 = 0.62 \text{ A} \tag{21}$$

Result

Electric current through $R_1=0.62 \text{ A}$

Voltage across $R_1=9.3 \text{ V}$

Correct answer

- E. 0.62 A and 9.3 V

For Question 1 in the multiple-choice format, ChatGPT's response is correct. ChatGPT accurately identified the parallel circuit concept, recognizing that the voltage across each branch equals the source voltage. Accordingly, the voltage across resistor R_1 was correctly determined as 9.3 V, and the current through R_1 was calculated using Ohm's law. This result demonstrates that ChatGPT applied the appropriate physical principles and produced an accurate solution for the given multiple-choice question.

Question 2 (Multiple-Choice)

Two resistors, $R_1=6 \Omega$ and $R_2=3 \Omega$, are connected in parallel. This parallel combination is then connected in

series with a third resistor, $R_3=8 \Omega$. When the circuit is connected to a 12 V voltage source, the electric current through and the voltage across R_1 are, respectively:

- A. 6 A and 12 V
- B. 4.5 A and 6 V
- C. 1.5 A and 4.5 V
- D. 0.5 A and 3 V
- E. 0.4 A and 2.4 V

ChatGPT's Response

Step 1: Equivalent resistance of the parallel combination

$$\frac{1}{R_{parallel}} = \frac{1}{R_1} + \frac{1}{R_2}$$

$$R_{parallel} = 2\Omega \tag{22}$$

Step 2: Total resistance of the circuit

$$R_{total} = 2\Omega + 8\Omega$$

$$R_{total} = 10\Omega \tag{23}$$

Step 3: Total current in the circuit

$$I_{total} = \frac{V}{R_{total}} = \frac{12}{10} = 1.2 \text{ A} \tag{24}$$

This is the current flowing through R_3 and into the parallel branch.

Step 4: Voltage across the parallel combination

$$V_p = I_{total} \times R_{parallel} = 2.4 \text{ V} \tag{25}$$

Because R_1 and R_2 are in parallel:

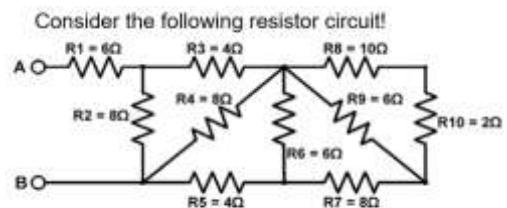
$$V_{R1} = V_p = 2.4 \text{ V} \tag{26}$$

Step 5: Current through R_1

$$I_{R1} = \frac{V_{R1}}{R_1} = \frac{2.4}{6} = 0.4 \text{ A} \tag{27}$$

For Question 2 in the multiple-choice format, ChatGPT's solution is correct and demonstrates procedural completeness. ChatGPT systematically simplified the parallel resistor combination, determined the total resistance of the series circuit, and accurately calculated the total current, the voltage across the parallel branch, and the current through resistor R_1 . The step-by-step presentation of intermediate calculations indicates an appropriate application of physical principles and a clear, coherent problem-solving process.

Question 3 (Multiple-Choice)



Consider the following resistor circuit!

The equivalent resistance between points A and B is ... Ω .

- A. 10 Ω
- B. 9 Ω
- C. 8 Ω
- D. 7.5 Ω
- E. 6.4 Ω

Figure 5. Question number 3

ChatGPT's Response

Step 1: Identify obvious series and parallel parts

Focus on R_8 , R_9 , R_{10} , and R_7 (Right-hand block):

R_8 (10Ω) is in series with R_{10} (2Ω)

$$R_{s1} = 10 + 2 = 12 \Omega \tag{28}$$

This result is in parallel with R_9 (6Ω)

$$\frac{1}{R} = \frac{1}{12} + \frac{1}{6}$$

$$R_{p1} = 4 \Omega \tag{29}$$

The result is then in series with R_7 (8Ω)

$$R_{s2} = 4 + 8 = 12 \Omega \tag{30}$$

Equivalent resistance of the right block = 12Ω .

Step 2: Combine with the middle branch

Between the upper-middle and lower-middle nodes:

$$R_6 = 6 \Omega$$

Right block = 12Ω

$$\frac{1}{R} = \frac{1}{6} + \frac{1}{12}$$

$$R_{p2} = 4 \Omega \tag{31}$$

Middle branch equivalent = 4Ω

Step 3: Combine with R_4 and R_5

Now look at the lower-left triangular section:

$R_5 = 4 \Omega$ is in series with the middle branch (4Ω):

$$R_{s3} = 4 + 4 = 8 \Omega \tag{32}$$

This result is in parallel with $R_4 = 8 \Omega$:

$$\frac{1}{R} = \frac{1}{8} + \frac{1}{8}$$

$$R_{p3} = 4 \Omega \tag{33}$$

Lower-left block equivalent = 4Ω

Step 4: Combine with R_3 and R_2

$R_3 = 4 \Omega$ is in series with the lower-left block (4Ω):

$$R_{s4} = 4 + 4 = 8 \Omega \tag{34}$$

This result is in parallel with $R_2 = 8 \Omega$:

$$\frac{1}{R} = \frac{1}{8} + \frac{1}{8}$$

$$R_{p4} = 4 \Omega \tag{35}$$

Middle section equivalent = 4Ω

Step 5: Final combination with R_1

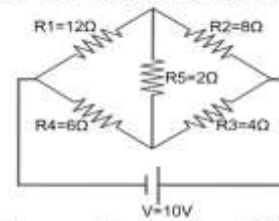
$R_1 = 6 \Omega$ is in series with the remaining equivalent resistance:

$$R_{\text{equivalent}} = 6 + 4 = 10 \Omega.$$

The solution generated by ChatGPT for this multiple-choice question is correct, systematic, and well-structured. ChatGPT applied a block-based simplification strategy, starting from the right-hand subcircuit and progressively reducing the circuit into simpler equivalent resistances. The selection of subcircuits to be simplified at each stage was appropriate, and the series-parallel relationships were consistently identified correctly. This structured approach enabled ChatGPT to manage the circuit complexity effectively and arrive at the correct final equivalent resistance.

Question 4 (Multiple-Choice)

Consider the following resistor circuit!



The current flowing through R_5 is ... A.
 A. 0
 B. 2
 C. 3
 D. 4
 E. 5

Figure 6. Question number 4

ChatGPT's Response

Step 1: Observe the bridge structure (key idea)

This circuit forms a Wheatstone bridge-like network.

Left side:

$$\frac{R_1}{R_4} = \frac{12}{6} = 2$$

Right side:

$$\frac{R_2}{R_3} = \frac{8}{4} = 2$$

The ratios are equal.

Step 2: Consequence of equal ratios

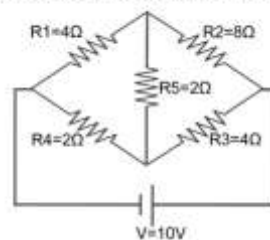
Because the bridge is balanced. That means:

The electric potential at the top node and bottom node of the bridge is the same. Therefore, no voltage difference exists across R_5 . $I_{R5} = 0$ A.

The solution provided by ChatGPT for this multiple-choice question is correct. ChatGPT successfully identified the circuit as a Wheatstone bridge and applied the appropriate balance condition by comparing the resistance ratios on both sides of the bridge. This correct recognition allowed ChatGPT to directly conclude that no current flows through resistor R_5 , demonstrating accurate conceptual understanding and efficient problem-solving.

Question 5 (Multiple-Choice)

Observe the Wheatstone bridge circuit below!



The current flowing through R_5 is ... A.
 A. 0
 B. 4
 C. 5
 D. 9
 E. 12

Figure 7. Question number 5

ChatGPT's Response

Step 1: Check the Wheatstone bridge balance condition.

Left side:

$$\frac{R_1}{R_4} = \frac{4}{2} = 2$$

Right side:

$$\frac{R_2}{R_3} = \frac{8}{4} = 2$$

The ratios are equal.

Step 2: Consequence of equal ratios

Because the bridge is balanced. That means:

The electric potential at the top node and the bottom node of the bridge is the same. Therefore, no voltage difference exists across R_5 . $I_{R5} = 0$ A.

The solution provided by ChatGPT for this multiple-choice question is correct. ChatGPT accurately identified the circuit as a Wheatstone bridge and correctly applied the balance condition by comparing the resistance ratios on both sides of the bridge. As a result, ChatGPT appropriately concluded that no current flows through resistor R_5 .

Comparative Discussion: Essay vs Multiple-Choice

The comparative analysis indicates that question format substantially influences AI performance in solving electrical circuit problems, despite identical content and cognitive demands. Differences between essay and multiple-choice formats were evident in accuracy, reasoning consistency, and conceptual reliability of ChatGPT's responses, supporting prior findings that assessment format plays a key role in shaping generative AI outputs (Xia et al., 2024).

Overall, ChatGPT demonstrated more stable and accurate performance in the multiple-choice format. All validated multiple-choice items were answered correctly, including problems involving complex series-parallel networks and Wheatstone bridge circuits. The presence of predefined answer options appears to constrain the AI's reasoning space, reducing conceptual ambiguity and guiding solution pathways toward correct physical principles. In this sense, multiple-choice items function not only as assessment tools but also as implicit scaffolding for AI reasoning, particularly in structured problem contexts.

In contrast, the essay format revealed greater variability in performance. While ChatGPT successfully solved several text-based essay questions, errors emerged primarily in image-based and structurally complex circuits. These errors were predominantly conceptual rather than computational and occurred at the initial stage of circuit interpretation, such as misidentifying series-parallel relationships or failing to recognize standard circuit configurations. Consequently, incorrect conceptual modeling led to invalid solution paths despite mathematically correct procedures. This finding extends earlier evidence

showing that AI can solve image-based physics problems when circuit configurations are relatively simple, but encounters difficulties as structural complexity increases (Nurhuda et al., 2023).

Differences were also observed in the transparency of reasoning. In essay responses, ChatGPT occasionally produced correct final answers without explicitly presenting intermediate steps, limiting the explanatory and instructional value of the solution. Conversely, multiple-choice responses consistently exhibited structured, sequential reasoning, suggesting that evaluating answer options encourages step-by-step verification before selecting a final response. Similar tendencies of generative AI to prioritize outputs over explicit reasoning processes have been noted in previous studies (Özçelik & Ekşi, 2024).

Prompt explicitness further influenced performance in the essay format. Essay questions that explicitly identified circuit types resulted in correct responses, whereas similar problems without such cues produced conceptual errors. In the multiple-choice format, correct pattern recognition occurred even without explicit instructions, indicating that the answer options themselves may serve as contextual signals that support circuit identification.

Taken together, these findings demonstrate that question format is a critical factor in assessing AI problem-solving performance in physics. Essay questions are more effective for exposing limitations in conceptual understanding, visual interpretation, and independent reasoning, whereas multiple-choice questions tend to yield higher apparent performance by constraining possible responses and reducing interpretive uncertainty. Reliance on multiple-choice items alone may therefore overestimate AI competence.

From an assessment perspective, essay-based tasks are better suited to evaluating the depth and robustness of AI reasoning, while multiple-choice tasks are better suited to measuring procedural accuracy and principle recognition. Using both formats is recommended to obtain a more comprehensive and balanced evaluation of AI performance in complex physics domains, such as electrical circuits.

Conclusion

This study demonstrates that question format significantly influences the evaluation of artificial intelligence performance in solving electrical circuit problems. Despite using identical content and cognitive demands, ChatGPT showed markedly different outcomes, achieving 60% accuracy on essay items and 100% on multiple-choice questions. Errors in essay responses were mainly conceptual, arising during initial problem interpretation, especially for image-based and

complex circuit tasks, whereas multiple-choice items showed no errors due to the guidance provided by predefined options. These results indicate that multiple-choice formats may overestimate AI capability by reducing interpretive complexity, while essay formats more effectively assess conceptual understanding and independent reasoning. Therefore, combining both formats is recommended to obtain a more comprehensive evaluation of AI performance, capturing both procedural accuracy and conceptual depth in physics problem solving.

Acknowledgments

The authors would like to thank SMA Negeri 4 Ternate for granting permission to conduct the data collection. The authors also express their appreciation to the expert validators for their valuable feedback during the instrument development and validation process.

Author Contributions

Conceptualization, A.S.; methodology, A.S.; software, A.S.; validation, A.S., M., and H.H.I.; formal analysis, A.S.; investigation, A.S.; resources, A.S.; data curation, A.S.; writing – original draft preparation, A.S.; writing – review and editing, M. and H.H.I.; visualization, A.S.; supervision, M. and H.H.I.; project administration, A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by PKUPT at the Faculty of Teacher Training and Education, Universitas Khairun, in 2025. The Article Processing Charge (APC) was funded by PKUPT FKIP Universitas Khairun.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Agyare, B., Asare, J., Kraishan, A., Nkrumah, I., & Adjekum, D. K. (2025). A cross-national assessment of artificial intelligence (AI) Chatbot user perceptions in collegiate physics education. *Computers and Education: Artificial Intelligence*, 8, 100365. <https://doi.org/10.1016/j.caeai.2025.100365>
- Aji, D. P., Akhsan, H., & Marlina, L. (2025). Development of Augmented Reality-Based and Artificial Intelligence-Assisted E-Modules on Global Warming Materials to Improve Critical Thinking Skills of High School Learners. *JPPIPA: Jurnal Penelitian Pendidikan IPA*, 11(2), 696–707. <https://doi.org/10.29303/jppipa.v11i2.9993>
- Ayuni, R. T., Jaedun, A., Zafrullah, Z., & Ramadhani, A. M. (2024). Trends in the Use of Artificial Intelligence in Science Education: Bibliometric & Biblioshiny Analysis (1975-2024). *JPPIPA: Jurnal Penelitian Pendidikan IPA*, 10(10), 740–756. <https://doi.org/10.29303/jppipa.v10i10.7846>
- Azap, S. (2025). The role of artificial intelligence in language teacher education: EFL learners' views on the concept of "ChatGPT" through metaphorical analysis. *Social Sciences & Humanities Open*, 11, 101394. <https://doi.org/10.1016/j.ssaho.2025.101394>
- Bates, T., Cobo, C., Mariño, O., & Wheeler, S. (2020). Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education*, 17(42), 1–12. <https://doi.org/10.1186/s41239-020-00218-x>
- Berliana, J. I., Winantoro, R. C., & Damariswara, R. (2024). Analisis Penggunaan AI dalam Mengerjakan Tugas pada Mahasiswa Universitas Nusantara PGRI Kediri Prodi PGSD Kelas 1C. *Jurnal Review Pendidikan Dan Pengajaran (JRPP)*, 7(3), 9724–9731. <https://doi.org/10.31004/jrpp.v7i3.30459>
- Cahyaningrum, I. Y., Fuady, A., & Sunismi, S. (2023). Analisis Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran Matematika Kelas VII dengan Berbantuan Aplikasi Software Anates. *Mathema Journal*, 5(2), 67–81. <https://doi.org/10.33365/jm.v5i2.2601>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(22), 1–22. <https://doi.org/10.1186/s41239-023-00392-8>
- Firman, F., & Firdaus, F. (2023). Efektivitas Artificial Intelligence (AI) dalam Menjawab Soal Higher Order Thinking Skill (HOTS). *Indonesian Journal of Educational Science (IJES)*, 6(1), 15–30. <https://doi.org/10.31605/ijes.v6i1.2899>
- Hariadi, A. F., Degeng, I. N. S., Setyosari, P., Ulfa, S., & Alhakim, F. F. (2025). The Effect of Hybrid Learning Method Assisted by Artificial Intelligence Camera Controller and Student Engagement on Learning Outcomes. *JPPIPA: Jurnal Penelitian Pendidikan IPA*, 11(9), 777–786. <https://doi.org/10.29303/jppipa.v11i9.12034>
- Lin, C., Huang, A. Y. Q., & Lu, O. H. T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments*, 10(41), 1–22. <https://doi.org/10.1186/s40561-023-00260-y>
- Mutakin, D. (2023). Pemanfaatan Software SPSS untuk Analisis Instrumen Penelitian dalam Penelitian Tindakan Kelas. *Jurnal Inovasi Dan Manajemen*

- Pendidikan*, 3(1), 33-49.
<https://doi.org/10.12928/jimp.v3i1.7727>
- Nam, B. H., & Bai, Q. (2023). ChatGPT and its ethical implications for STEM research and higher education: A media discourse analysis. *International Journal of STEM Education*, 10(66), 1-24. <https://doi.org/10.1186/s40594-023-00452-5>
- Niu, Z., Tan, Z.-M., Yu, H., Gu, J.-F., Chen, G., & Huang, W. (2025). Typhoon Science Meets Artificial Intelligence: A Roundtable on Bridging Physics Based and Data-Driven Paradigms. *Tropical Cyclone Research and Review*, 14(3), 311-316. <https://doi.org/10.1016/j.tcr.2025.08.006>
- Nurhuda, D., Kumala, S. A., & Widiyatun, F. (2023). Analisis Kecerdasan Buatan ChatGPT dalam Penyelesaian Soal Fisika Bergambar pada Materi Resistor. *Jurnal Luminous: Riset Ilmiah Pendidikan Fisika*, 4(2), 62-70. <https://doi.org/10.31851/luminous.v4i2.12232>
- Özçelik, N. P., & Ekşi, G. Y. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant – A case study. *Smart Learning Environments*, 11(10), 1-18. <https://doi.org/10.1186/s40561-024-00296-8>
- Pairunan, A. S., Nurjannah, N., Haeruddin, H., Wahyono, U., & Pasaribu, M. (2025). Development of Artificial Intelligence Based E-Learning Modules to Improve Students' Critical Thinking and Communication Skills. *JPPIPA: Jurnal Penelitian Pendidikan IPA*, 11(9), 993-1005. <https://doi.org/10.29303/jppipa.v11i9.12050>
- Prananta, A. W., Susanto, N., Purwantoro, A., & Fuadah, N. (2023). ChatGPT Artificial Intelligence Integration in Science Learning Media: Systematic Literature Review. *JPPIPA: Jurnal Penelitian Pendidikan IPA*, 9(7), 315-321. <https://doi.org/10.29303/jppipa.v9i7.4386>
- Prasetya, F., Fortuna, A., Samala, A. D., Latifa, D. K., Andriani, W., Gusti, U. A., ... García, J. L. C. (2025). Harnessing artificial intelligence to revolutionize vocational education: Emerging trends, challenges, and contributions to SDGs 2030. *Social Sciences & Humanities Open*, 11, 101401. <https://doi.org/10.1016/j.ssaho.2025.101401>
- Purba, S., Waluyo, B. D., Pangaribuan, W., Zulkarnain, S. A. B., & Astrid, E. (2025). Development and Implementation of AI-Driven Learning Media for Low-Power Inverter Mastery Using the ADDIE Model. *JPPIPA: Jurnal Penelitian Pendidikan IPA*, 11(8), 1098-1109. <https://doi.org/10.29303/jppipa.v11i8.11573>
- Purnama, I., Wijaya, R. F., Harahap, A., & Edi, F. (2025). Artificial Intelligence as a Science Teacher Assistant: An Analysis of Machine Learning Utilization in Diagnosing Student Misconceptions: A Review. *JPPIPA: Jurnal Penelitian Pendidikan IPA*, 11(12), 1-7. <https://doi.org/10.29303/jppipa.v11i12.13089>
- Puspitasari, W. D., & Febrinita, F. (2021). Penguujian Validasi Isi (Content Validity) Angket Persepsi Mahasiswa terhadap Pembelajaran Daring Matakuliah Matematika Komputasi. *Journal Focus Action of Research Mathematic (Factor M)*, 4(1), 77-90. https://doi.org/10.30762/factor_m.v4i1.3254
- Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 19(21), 1-20. <https://doi.org/10.1186/s41239-022-00326-w>
- Schmidt, D. A., Alboloushi, B., Thomas, A., & Magalhaes, R. (2025). Integrating artificial intelligence in higher education: Perceptions, challenges, and strategies for academic innovation. *Computers and Education Open*, 9, 100274. <https://doi.org/10.1016/j.caeo.2025.100274>
- Shahzad, M. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: The role of trust as a cornerstone. *International Journal of Educational Technology in Higher Education*, 21(46), 1-23. <https://doi.org/10.1186/s41239-024-00478-x>
- Smith-Mutegi, D., Mamo, Y., Kim, J., Crompton, H., & McConnell, M. (2025). Perceptions of STEM education and artificial intelligence: A Twitter (X) sentiment analysis. *International Journal of STEM Education*, 12(9), 1-18. <https://doi.org/10.1186/s40594-025-00527-5>
- Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*, 6, 100169. <https://doi.org/10.1016/j.caeo.2024.100169>
- Sugiyono, S. (2021). *Metode Penelitian Kualitatif*. Bandung: Alfabeta.
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems With Applications*, 252, 124167. <https://doi.org/10.1016/j.eswa.2024.124167>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(40), 1-22. <https://doi.org/10.1186/s41239-024-00468-z>
- Yusof, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, 11(2), 49-54. <https://doi.org/10.21315/eimj2019.11.2.6>

- Zawacki-Richter, O., Bai, J. Y. H., Lee, K., Tryon, P. J. S. van, & Prinsloo, P. (2024). New advances in artificial intelligence applications in higher education? *International Journal of Educational Technology in Higher Education*, 21(32), 1-4. <https://doi.org/10.1186/s41239-024-00464-3>
- Zulpan, Z., Sahrul, S., Yusron, A., Hanida, R. S., & Marfu'ah, S. (2024). Validitas Tes Objektif Bentuk Pilihan Ganda. *Jurnal Pendidikan Matematika*, 1(3), 1-10. <https://doi.org/10.47134/ppm.v1i3.812>