



Comparing Accuracy, Language, and Pedagogy in AI-Generated Exponential Equation Solutions

Yusuf Adhitya^{1*}, Bill Chairy Rizki Bustaren², Jay Fie Paler Luzano³

^{1,2}Mathematics Education, Indonesia University of Education, Bandung, Indonesia

³College of Education, Bukidnon State University, Philippines

*Correspondence: E-mail: yusufadhitya@upi.edu

A B S T R A K	A R T I C L E I N F O
<p><i>Pertumbuhan AI generatif telah mengubah cara belajar matematika dengan menyediakan solusi instan untuk berbagai masalah. Studi ini menilai empat model AI populer, yakni MathGPT, ChatGPT, DeepSeek, dan Gemini AI, pada tiga soal persamaan eksponensial. Penilaian dilakukan berdasarkan tiga aspek: akurasi, bahasa, dan pedagogi. Akurasi memastikan jawaban benar, konsisten, dan diverifikasi secara langsung. Bahasa menilai gaya narasi dalam langkah prosedural, penggunaan notasi yang tepat, serta terminologi yang formal. Pedagogi memeriksa komponen scaffolding, cara mengomunikasikannya, serta kesesuaiannya dengan berbagai tipe pembelajar. Hasilnya menunjukkan bahwa keempat AI tersebut mampu memberikan jawaban yang benar, meski penjelasannya berbeda-beda. DeepSeek menjelaskan secara rinci; MathGPT unggul dalam penggunaan notasi formal dan terminologi; ChatGPT memberikan solusi ringkas dan langkah demi langkah, sementara Gemini AI menyusun scaffolding naratif yang cerdas. Temuan ini mengindikasikan bahwa hasil jawaban AI tidak hanya bergantung pada kebenaran jawaban, tetapi juga pada cara penjelasan disampaikan dan kecocokannya dengan kebutuhan pembelajar. Pendidik dan pengembang dapat memanfaatkan keunggulan masing-masing AI untuk menciptakan alat pembelajaran yang lebih adaptif dan tepat sasaran.</i></p>	<p>Article History: Received: 2026-06-08 Revision: 2026-06-10 Accepted: 2026-06-11 Published: 2026-06-11</p> <p>Kata Kunci: ChatGPT MathGPT DeepSeek Gemini AI Generatif Persamaan Eksponensial</p>
<p>A B S T R A C T</p> <p><i>The growth of generative AI has transformed how mathematics is learned by providing instant solutions to a wide range of problems. This study evaluates four popular AI models, namely MathGPT, ChatGPT, DeepSeek, and Gemini AI, on three exponential equation problems. The assessment is based on three aspects: accuracy, language, and pedagogy. Accuracy ensures the answers are correct, consistent, and directly verified. Language assesses the narrative style in procedural steps, the use of appropriate</i></p>	<p>ChatGPT MathGPT DeepSeek Gemini Generative AI Exponential Equation</p>

notation, and the use of formal terminology. Pedagogy examines the scaffolding components, how they are communicated, and their suitability for different types of learners. The results show that all four AI models can provide correct answers, although their explanations vary. DeepSeek explains in detail. MathGPT excels at using formal notation and terminology; ChatGPT offers concise solutions with step-by-step guidance, while Gemini AI constructs intelligent narrative scaffolding. These findings indicate that AI answer quality depends not only on correctness but also on how the explanations are delivered and their suitability for learners' needs. Educators and developers can leverage each AI's strengths to create more adaptive and targeted learning tools.

© 2026 Kantor Jurnal dan Publikasi UPI

1. INTRODUCTION

The emergence of generative artificial intelligence (AI) has significantly transformed the educational landscape by enabling learners to instantly access explanations, examples, and solutions across various disciplines. Unlike traditional search engines, AI can generate step-by-step solutions, deliver natural-language explanations, and respond interactively to follow-up questions. In mathematics education, AI has the potential to enhance students' understanding and improve their performance, especially in collaborative learning environments (Canonigo, 2024; Hwang, 2022). Consequently, many students now depend on AI not only to obtain answers but also to grasp mathematical concepts and procedures. Additionally, AI can personalize teaching methods and expand access to digital resources, providing real-time, consistent feedback that supports learning (Copur-Gencturk et al., 2024; Opesemowo & Ndlovu, 2024; Zhang & Zhang, 2024). Overall, AI serves as a valuable partner for both teachers and students, advancing and enriching mathematics education.

Although the development of AI provides many benefits, concerns about its usefulness are also numerous. Issues related to AI usually concern ensuring AI reliability (Annua et al., 2025), as it can give incorrect answers (Opesemowo & Ndlovu, 2024), exhibit inconsistent reasoning or bias (Wardat et al., 2023), or serve as a companion in the use of AI (Maulida et al., 2024). Particularly in mathematics education, it is crucial to understand concepts, engage in logical reasoning, communicate clearly, and solve problems. Focusing solely on answer correctness may overlook a deep understanding of the concept (Arifin et al., 2025). Students might also place excessive trust in AI solutions, which can ultimately decrease critical thinking, creativity, and independent learning (Rahmatika et al., 2025). Therefore, AI should be evaluated not just for accuracy but also for its ability to communicate mathematical ideas effectively and promote meaningful learning.

Despite the growing body of research on generative AI in education, several gaps remain. Many studies, such as Islam et al. (2026), Macina et al. (2025), and Xie et al. (2024), primarily focus on the accuracy of AI-generated answers, often reporting success rates on mathematical tasks while paying limited attention to the quality of explanations. The correct answer regarding AI has not yet been studied pedagogically, whether it is accurate in notation or in its presentation. Additionally, the above studies remain limited to a single type of AI rather than comparisons between different AIs. Generally, there is still a limited understanding of how different AI systems perform simultaneously across mathematical accuracy, language characteristics, and pedagogical support.

This study aims to compare the performance of four popular generative AIs (MathGPT, ChatGPT, DeepSeek, and Gemini) in solving exponential equation problems. The four AIs were chosen because they are similar: all are generative, equally popular, widely used, and offer free versions. These problems are significant in secondary and undergraduate mathematics, involving exponents, algebra, and reasoning. The research introduces a holistic evaluation framework that considers not only correctness but also accuracy, language clarity, formalism, and pedagogical support, such as explanations and guidance. By integrating these dimensions, the study provides a comprehensive view of AI's educational value, offering insights for educators, researchers, and students on the effective and critical use of generative AI as a learning aid in mathematics.

2. METHOD

The method used is a descriptive qualitative research approach. The data source in this study is the output of four generative AIs solving exponential equation problems. The four AIs are MathGPT, ChatGPT, DeepSeek, and Gemini AI. Because DeepSeek and Gemini AI are free, the researchers also used the other two AIs, namely MathGPT and ChatGPT, in free versions. This is important because this research will focus on the fundamental aspects of the

four AIs' features. The use of paid AIs is feared to introduce bias, potentially making this study unbalanced.

The four AIs are given the same prompt or command, such as “*please solve for x for ...*,” to ensure fairness in the analysis process. Only the first response from each AI is collected and analyzed to represent the initial answer a user would most likely receive in everyday situations. The three questions used in this study are mathematical problems involving exponential equations. They are arranged from easiest to most difficult, in order of increasing complexity. Additionally, the questions are ordered according to the specific skills needed to solve exponential equations. The complete list of questions examined in this study is presented in Table 1 below.

Table 1. Problem of Exponential Equation

Problem	Equation	Complexity	Key Skill Tested
1	$3^{2x-1} = \frac{1}{27}$	Basic	Base conversion, substitution
2	$4^{x^2-x-4} = \left(\frac{1}{2}\right)^{2-2x}$	Intermediate	Quadratic transformation
3	$2^{2x+1} - 9 \cdot 2^x + 4 = 0$	Advanced	Substitution + quadratic formula

The analysis focused on three main aspects: the accuracy of the answers, the use of language, and the pedagogical approach. The following is an explanation of each aspect as seen in Table 2.

Table 2. Definition of The Aspects

No	Aspect	Definition
1	Accuracy	The accuracy and consistency of logic in solving problems, including whether a verification step is present
2	Language	The level of formality, the accuracy of mathematical notation, and the clarity of explanation
3	Pedagogy	The existence of scaffolding steps, communication styles, and the extent to which AI confirms the correctness of the answer

3. RESULT AND DISCUSSION

This section begins by discussing the findings of each problem. It then emphasizes integrating these findings based on accuracy and pedagogy. To aid discussion, each AI is assigned a specific color: MathGPT (red), ChatGPT (blue), DeepSeek (yellow), and Gemini AI (green).

Problem 1

Generally, Problem 1 asked for the value of x in the equation. $3^{2x-1} = \frac{1}{27}$. The problem can be solved by matching the base of the exponential equation. In such problems, the easiest step is to change to base 3. So, the first important step is to change $\frac{1}{27}$ to 3^{-3} . Because the base is the same, the exponential equation can be solved by equalizing its exponent to obtain a linear

equation of a variable (SPLSV) of $2x - 1 = -3$. Lastly, using algebraic calculations, it obtains $x = -1$.

With the same prompt, the results of the four AIs are shown in Figure 1 below.

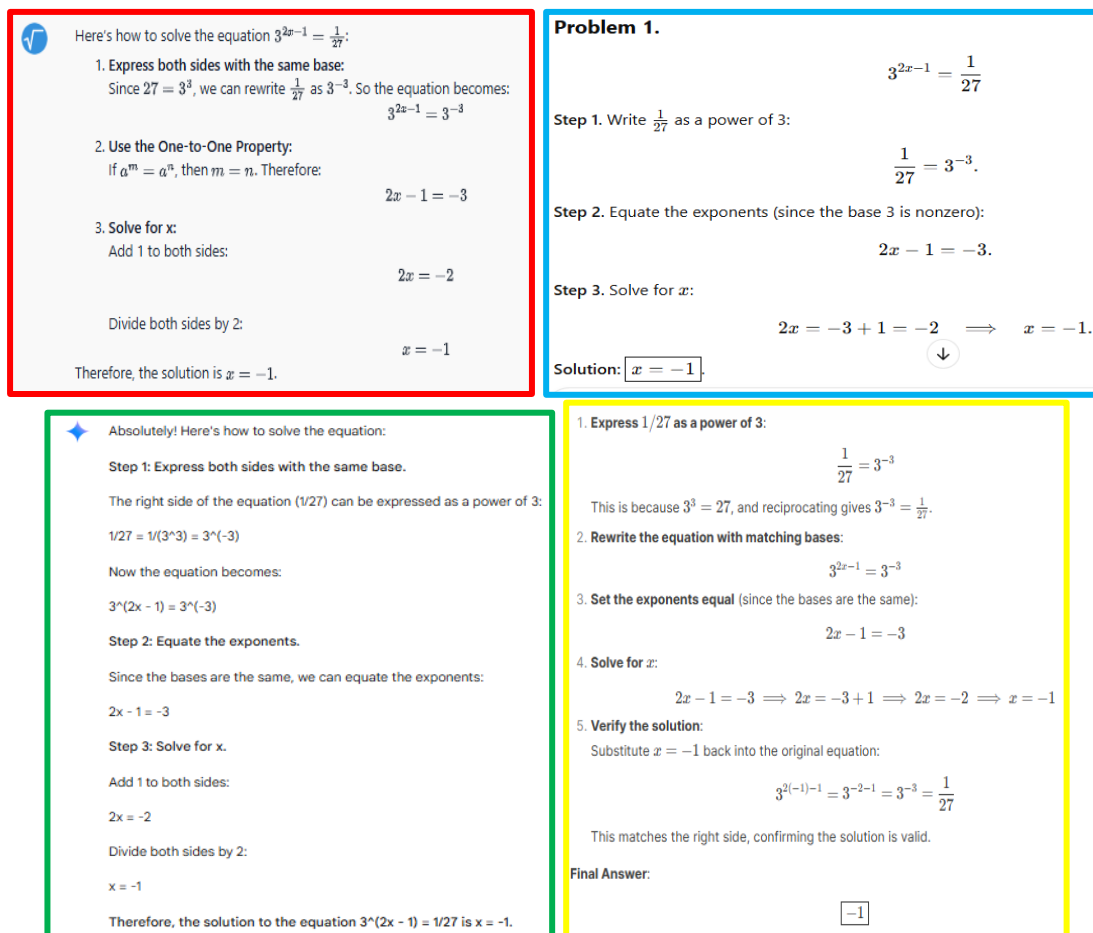


Figure 1. AI Outputs in Solving Problem 1

In the first problem, all AIs produce the correct answer, namely $x = -1$, by changing the base to 3 and equalizing their exponents. However, in terms of pedagogy, there are 2 AIs (MathGPT and Gemini) that use formal teaching languages. This is evident in the output of the word 'we,' which is impressive; AI considers us as a partner in solving this problem. Unlike ChatGPT and DeepSeek, they only give a procedural list. In addition, the pedagogy aspect also observes that only MathGPT uses formal terms such as "one-to-one properties, that is, if $a^m = a^n$ then $m = n$. The term provides mathematical rigor that may not be understood by beginners (non-mathematicians). Additionally, it does not display verification steps to support its answers. One positive aspect of this AI is that it solves linear equations clearly, leveraging the principle of similarity of treatment (adding 1 to both sides of the equation). It is beneficial for beginners to understand linear equations in a meaningful way. The second AI, ChatGPT, solves the problem with a simple step, but omits a verification step. ChatGPT explains why three is non-zero to emphasize that exponential equations apply. Like previous AI, ChatGPT also does not provide a verification step. DeepSeek offers a perfect solution by adding verification stages. It confirms that the solution is valid. The last AI, Gemini, explains in simple and detailed language, even without verification stages. The weakness of Gemini is that its output does not support mathematical text, making it less comfortable to read. To facilitate comparison of the steps carried out by the four AIs, please refer to Table 4 below.

Problem 2

Problem 2 asked for the value of x in the equation $4^{x^2-x-4} = \left(\frac{1}{2}\right)^{2-2x}$. The problem can be solved by setting both bases of the exponential equation equal. One alternative is to change the base to 2, so it makes $4 = 2^2$ and $\frac{1}{2} = 2^{-1}$. Because the base is the same, the exponential equation can be solved by equating its exponent $2(x^2 - x - 4) = -1(2 - 2x)$. By the algebraic process, it will form a quadratic equation, namely $x^2 - 2x + 3 = 0$. Furthermore, the equation will have the roots of $x = 3$ and $x = -1$. Using the prompt, “Please solve for x in $4^{x^2-x-4} = \left(\frac{1}{2}\right)^{2-2x}$, Here is a breakdown of the four AIs’ output.

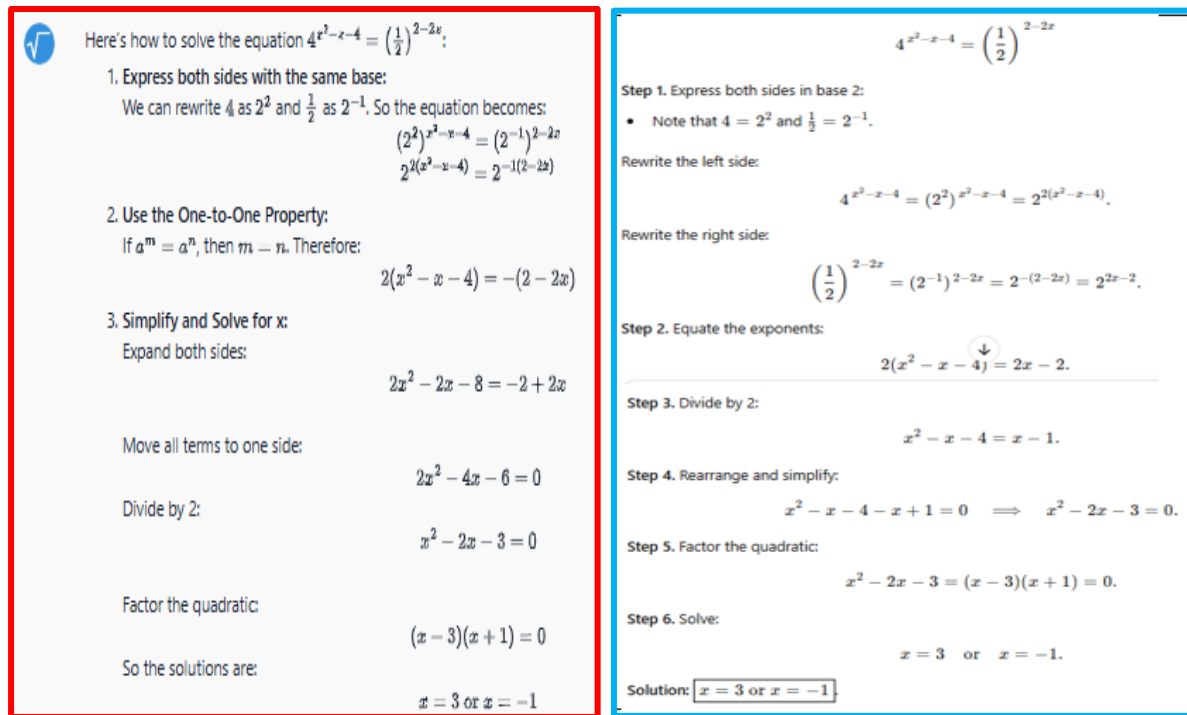


Figure 2. (a) MathGPT and ChatGPT in Solving Problem 2

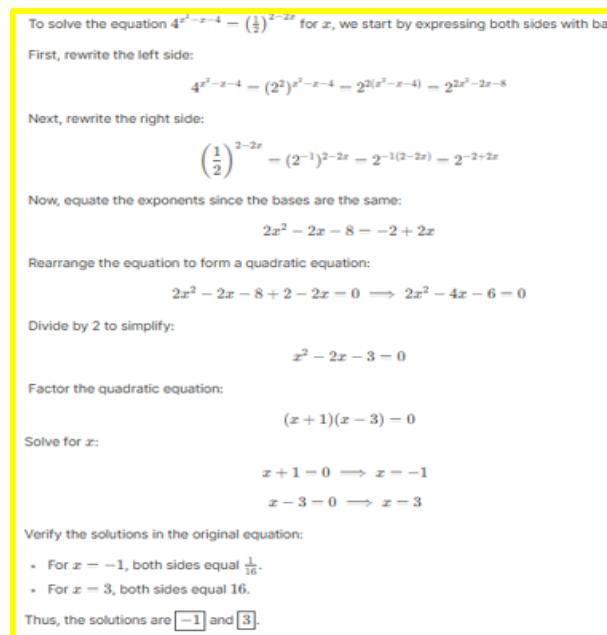


Figure 2. (b) DeepSeek in Solving Problem 2

$4^{(x^2 - x - 4)} = (1/2)^{(2 - 2x)}$ <p>Step 1: Express both sides with the same base.</p> <p>We can express both 4 and 1/2 as powers of 2:</p> <ul style="list-style-type: none"> • $4 = 2^2$ • $1/2 = 2^{-1}$ <p>Substitute these into the equation:</p> $(2^2)^{(x^2 - x - 4)} = (2^{-1})^{(2 - 2x)}$ <p>Step 2: Simplify using exponent rules.</p> <p>When raising a power to a power, multiply the exponents:</p> $2^{2(x^2 - x - 4)} = 2^{-1(2 - 2x)}$ <p>Simplify further:</p> $2^{(2x^2 - 2x - 8)} = 2^{(-2 + 2x)}$ <p>Step 3: Equate the exponents.</p> <p>Since the bases are the same, we can equate the exponents:</p> $2x^2 - 2x - 8 = -2 + 2x$	<p>Step 4: Rearrange into a quadratic equation.</p> <p>Move all terms to one side to get a quadratic equation:</p> $2x^2 - 2x - 8 + 2 - 2x = 0$ $2x^2 - 4x - 6 = 0$ <p>Step 5: Simplify the quadratic equation.</p> <p>Divide the entire equation by 2 to make it simpler:</p> $x^2 - 2x - 3 = 0$ <p>Step 6: Solve the quadratic equation.</p> <p>We can solve this by factoring:</p> $(x - 3)(x + 1) = 0$ <p>This gives us two possible solutions:</p> $x - 3 = 0 \Rightarrow x = 3$ $x + 1 = 0 \Rightarrow x = -1$ <p>Therefore, the solutions to the equation are $x = 3$ and $x = -1$.</p>
--	--

Figure 2. (c) Gemini in Solving Problem 2

In general, all four AIs answered Problem 2 using the same process, with slight variations in the final answers. All the AIs convert the base to 2 and then equalize the exponents. The algebraic process is also similar, beginning with forming a quadratic equation and then factoring to find its roots. The roots of the quadratic equation are the solutions to the exponential equation.

There is a slight difference in ChatGPT, where it immediately simplifies $2(x^2 - x - 4) = 2x - 2$ to $x^2 - x - 4 = x - 1$ without first expanding on the left side. Other AIs expand first, then divide them by two. Notably, MathGPT consistently mentions the term "one-to-one properties" in its solution output. DeepSeek demonstrates consistency by providing verification steps, and Gemini explains each step in a detailed and reasonable manner, but continues to use non-mathematical language. Unfortunately, even though Deepseek uses a verification step, only ChatGPT writes the final answer correctly $x = 3$ or $x = -1$. On the other hand, the other three AIs use "and" instead of "or". This is a conceptual error in solving quadratic equations. According to MathGPT, it is supposed to use the word "or" in the process, but in the final answer, it changes to "and". This is quite unusual and demonstrates that AI can still be inconsistent in providing answers.

Problem 3

Problem 3 is an advanced problem with the form $Af(x)^2 + Bf(x) + C = 0$ with $f(x)$ is an exponential number. In Problem 3, $f(x)$ is 2^x so that the equation formed is $2f(x)^2 - 9f(x) + 4 = 0$. Writing of $f(x)$ is often replaced by other variables, such as y or p , to make writing easier. The value needs to be substituted back into the equation to find the final answer. In this problem, the value of $f(x)$ is $1/2$ or 2 , so the value of x is -1 or 2 .

Based on the AI process with the prompt "Please solve for x in $2^{2x+1} - 9 \cdot 2^x + 4 = 0$. The following results are obtained.

Here's how to solve the equation $2^{2x+1} - 9 \cdot 2^x + 4 = 0$:

- Rewrite the equation:**
We can rewrite 2^{2x+1} as $2 \cdot 2^{2x}$ or $2 \cdot (2^x)^2$. So the equation becomes:
 $2 \cdot (2^x)^2 - 9 \cdot 2^x + 4 = 0$
- Substitution:**
Let $y = 2^x$. Then the equation becomes:
 $2y^2 - 9y + 4 = 0$
- Solve the quadratic equation:**
We can factor the quadratic equation:
 $(2y - 1)(y - 4) = 0$
So the solutions for y are:
 $y = \frac{1}{2}$ or $y = 4$
- Substitute back:**
Now we substitute back 2^x for y :
 $2^x = \frac{1}{2}$ or $2^x = 4$
- Solve for x :**
For $2^x = \frac{1}{2}$, we can rewrite $\frac{1}{2}$ as 2^{-1} . So:
 $2^x = 2^{-1}$
 $x = -1$
For $2^x = 4$, we can rewrite 4 as 2^2 . So:
 $2^x = 2^2$
 $x = 2$
Therefore, the solutions are $x = -1$ and $x = 2$.

Figure 3. (a) MathGPT in Solving Problem 3

Step 1. Make the substitution:
 $t = 2^x$, with $t > 0$.

Then note:
 $2^{2x+1} = 2 \cdot 2^{2x} = 2t^2$.

The equation becomes:
 $2t^2 - 9t + 4 = 0$.

Step 2. Solve the quadratic equation:
 $2t^2 - 9t + 4 = 0$.

Calculate the discriminant:
 $\Delta = (-9)^2 - 4 \cdot 2 \cdot 4 = 81 - 32 = 49$.

Find the roots:
 $t = \frac{9 \pm 7}{4}$.

Thus,
 $t = \frac{16}{4} = 4$ or $t = \frac{2}{4} = \frac{1}{2}$.

Step 3. Substitute back $t = 2^x$:

- For $2^x = 4$:
 $4 = 2^2$ so $x = 2$.
- For $2^x = \frac{1}{2}$:
 $\frac{1}{2} = 2^{-1}$ so $x = -1$.

Solution: $x = 2$ or $x = -1$.

Figure 3. (b) ChatGPT in Solving Problem 3

To solve the equation $2^{2x+1} - 9 \cdot 2^x + 4 = 0$, follow these steps:

- Substitute $y = 2^x$:**
 - Let $y = 2^x$. Then $2^{2x} = (2^x)^2 = y^2$, and $2^{2x+1} = 2 \cdot y^2$.
 - The equation becomes:
 $2y^2 - 9y + 4 = 0$
- Solve the quadratic equation:**
 - Use the quadratic formula $y = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ with $a = 2$, $b = -9$, and $c = 4$:
 $y = \frac{9 \pm \sqrt{81 - 32}}{4} = \frac{9 \pm 7}{4}$
 - Solutions:
 $y = \frac{16}{4} = 4$ or $y = \frac{2}{4} = 0.5$
- Solve for x :**
 - For $y = 4$:
 $2^x = 4 \implies x = 2$ (since $4 = 2^2$)
 - For $y = 0.5$:
 $2^x = 0.5 \implies x = -1$ (since $0.5 = 2^{-1}$)
- Verify solutions:**
 - $x = 2$:
 $2^{2(2)+1} - 9 \cdot 2^2 + 4 = 32 - 36 + 4 = 0$
 - $x = -1$:
 $2^{2(-1)+1} - 9 \cdot 2^{-1} + 4 = 0.5 - 4.5 + 4 = 0$

Final Solutions: $x = -1$ and $x = 2$

Figure 3. (c) DeepSeek in Solving Problem 3

<p>◆ Absolutely! Let's solve the equation:</p> $2^{2x+1} - 9 \cdot 2^x + 4 = 0$ <p>Step 1: Simplify the first term.</p> <p>Using exponent rules, we can rewrite 2^{2x+1} as:</p> $2^{2x+1} = 2^{2x} \cdot 2^1 = 2 \cdot 2^{2x} = 2 \cdot (2^x)^2$ <p>Step 2: Substitute to create a quadratic equation.</p> <p>Let $y = 2^x$. Substitute this into the equation:</p> $2 \cdot y^2 - 9y + 4 = 0$ <p>Step 3: Solve the quadratic equation.</p> <p>We can solve this quadratic equation by factoring:</p> $(2y - 1)(y - 4) = 0$ <p>This gives us two possible solutions for y:</p>	<p>This gives us two possible solutions for y:</p> <ul style="list-style-type: none"> $2y - 1 = 0 \Rightarrow 2y = 1 \Rightarrow y = 1/2$ $y - 4 = 0 \Rightarrow y = 4$ <p>Step 4: Substitute back to find x.</p> <p>Remember that $y = 2^x$. Let's substitute back and solve for x:</p> <ul style="list-style-type: none"> Case 1: $y = 1/2$ $2^x = 1/2$ $2^x = 2^{-1}$ $x = -1$ Case 2: $y = 4$ $2^x = 4$ $2^x = 2^2$ $x = 2$ <p>Therefore, the solutions to the equation are $x = -1$ and $x = 2$.</p>
---	--

Figure 3. (d) Gemini in Solving Problem 3

Based on Figure 3, all AIs approach this problem with the same concept. The process involves substituting one variable for another. MathGPT, DeepSeek, and Gemini use the variable “y,” while ChatGPT uses the variable “t.” The second step is for all AIs to solve the quadratic equation and find its solutions. The final step is to substitute back and determine the value of x. Consistently, only DeepSeek employs a verification stage to check the accuracy of its solutions.

Unlike in Problem 2, where all AIs solve quadratic equations by factorization, the process in Problem 3 is different. MathGPT and Gemini use factorization. ChatGPT and DeepSeek use a quadratic formula. However, ChatGPT does not mention that this is a quadratic formula. It calculates the discriminant and then looks for the roots. Meanwhile, DeepSeek surprisingly provides a clear explanation regarding the quadratic formula $y = -\frac{b \pm \sqrt{b^2 - 4ac}}{2a}$. In this problem, DeepSeek and MathGPT are more mathematical in the process than other AIs.

The unfortunate aspect is that the conceptual error in the final answer occurs again. MathGPT remains inconsistent in its use of "or" throughout the process, switching to "and" in the final answer. DeepSeek and Gemini still use "and," which shows that both AIs share a similar conceptual approach. This is a recurring issue from the previous problem. ChatGPT is not as strong as MathGPT and DeepSeek in its mathematical reasoning, but the correct final answer indicates that this AI is also pedagogically reliable. Although Gemini tries to provide detailed explanations, it has a fundamental shortcoming. It cannot generate mathematical text, which is an important finding that warrants attention.

Summary of Findings Across Problems

Although each problem involved different forms of exponential equations, several recurring patterns emerged across the AI responses. To highlight these patterns, the results were summarized according to three analytical dimensions: accuracy, language, and pedagogy, as shown in Table 3.

Table 3. Summary of Findings

AI	Accuracy	Language	Pedagogical
MathGPT	Correct in all problems, but inconsistent use of 'and' instead of 'or' in Problems 2 and 3.	Uses formal math language and technical terms, such as the one-to-one property.	Highlights mathematical rigor and concepts, but may challenge beginners.
ChatGPT	Correct in all problems and consistently presents the final answer using 'or'.	Clear, concise, and procedural language.	Provides step-by-step solutions but lacks verification.
DeepSeek	Correct procedures in all problems, but inconsistent use of "and" in Problems 2 and 3.	Mathematical and structured language with clear formula explanations.	Includes verification steps that foster reflective thinking.
Gemini	Follows correct procedures, but inconsistent use of 'and' in Problems 2 and 3.	Simple, accessible language, but mathematical notation is less readable.	Provides basic explanations but lacks verification.

Table 3 highlights several recurring patterns across the three problems. Although all AI systems applied appropriate mathematical procedures, differences emerged in the consistency of final answers, the level of mathematical formality, and the pedagogical support provided. ChatGPT showed the most consistent accuracy, MathGPT emphasized mathematical rigor, DeepSeek consistently incorporated verification, and Gemini provided the most detailed and accessible explanations.

Discussion

The findings of this study indicate that the performance of AI in solving exponential equations should be evaluated through three complementary dimensions: accuracy, language, and pedagogy. From an accuracy perspective, all AI systems correctly identified the underlying mathematical concepts required to solve the three exponential-equation problems. They successfully transformed exponential expressions, applied properties of exponents, and used algebraic techniques such as factoring and substitution. This finding suggests that current AI systems possess strong procedural competence in handling routine exponential equation tasks. However, the analysis also revealed a recurring weakness. In Problems 2 and 3, MathGPT, DeepSeek, and Gemini incorrectly used the conjunction "and" instead of "or" when presenting multiple solutions. Although this error did not affect the numerical values obtained, it represents a conceptual inaccuracy because the solutions of an equation are alternative possibilities rather than simultaneous conditions. Moon's research (2025) shows that Gemini can solve procedural questions but lacks conceptual understanding. Interestingly, ChatGPT was the only AI that consistently provided the correct final answers across all three problems. This finding is consistent with the fact that the ChatGPT solution outperforms other AI solutions (Lie et al., 2024; Moon, 2025). Its findings also remind us that mathematical accuracy should be assessed not only by computational correctness but also by the precision of mathematical communication.

The second finding concerns differences in language use. Each AI exhibited a distinct communication style that reflects different priorities in explaining mathematics. MathGPT consistently employed formal mathematical terminology, as reported in Scarlatos & Lan (2023). In this case, MathGPT included concepts such as the one-to-one property of exponential functions. This approach strengthens mathematical rigor and exposes learners to

discipline-specific vocabulary. However, such terminology may be difficult for beginners lacking a sufficient mathematical background. In contrast, ChatGPT tended to use concise, procedural explanations, focusing on the essential steps to reach a solution. DeepSeek combined mathematical explanations with structured reasoning, while Gemini favored simple and accessible language that may be easier for novice learners to understand. These findings suggest that AI systems vary not only in what they explain but also in how they communicate mathematical ideas. Consequently, different AI tools may be suitable for different learner profiles and instructional purposes (Wardat et al., 2024).

The pedagogical analysis revealed perhaps the most meaningful distinction among the four AI systems. While all AIs provided step-by-step solutions, only DeepSeek consistently included verification stages to confirm the validity of the obtained answers. AI typically focuses only on the final solution (Hubert et al., 2025). Verification is an important component of mathematical thinking because it encourages learners to evaluate and justify solutions rather than merely follow procedures (Sa'adah et al., 2023). In contrast, MathGPT emphasized conceptual understanding through formal mathematical reasoning, ChatGPT focused on efficient problem-solving procedures, and Gemini provided detailed explanations intended to support comprehension. These differences reflect distinct pedagogical orientations: rigor-oriented (MathGPT), efficiency-oriented (ChatGPT), reflection-oriented (DeepSeek), and accessibility-oriented (Gemini). Therefore, pedagogical quality cannot be reduced to the length of an explanation or the correctness of an answer alone. It also involves how effectively the AI supports mathematical reasoning and learning processes.

Overall, the findings demonstrate that the strengths and weaknesses of AI-generated mathematical solutions extend beyond answer correctness. Accuracy, language, and pedagogy interact to shape the educational value of AI responses. An AI may provide a correct answer but lack pedagogical support, while another may offer detailed explanations yet contain subtle conceptual inconsistencies. These results highlight the importance of evaluating AI systems holistically when considering their use as learning companions in mathematics education.

4. CONCLUSION

This study found that all four AI (MathGPT, ChatGPT, DeepSeek, and Gemini) were generally capable of solving exponential equation problems correctly and applying appropriate mathematical procedures. However, differences emerged in terms of accuracy, consistency, language use, and pedagogical features. ChatGPT demonstrated the most consistent accuracy in presenting final answers, while MathGPT provided the highest level of mathematical rigor through formal terminology and conceptual explanations. DeepSeek distinguished itself by consistently including verification steps, whereas Gemini offered the most accessible and detailed explanations for beginners. Despite their overall strong performance, recurring inconsistencies in expressing multiple solutions, particularly the use of *and* instead of *or*, indicate that AI-generated responses still require critical evaluation. These findings suggest that the educational value of AI in mathematics extends beyond obtaining correct answers and should also consider the quality of explanation, mathematical communication, and pedagogical support provided to learners.

This study has limitations, mainly because it does not use a paid AI version and relies solely on the AI's initial response. Responses may vary if follow-up questions are asked, such as why AI chose a particular process or if other options are available. Despite these limitations, the researcher believes the findings can provide new insights for future research. Additional studies could explore issues like AI output bias, misconceptions, and differences between AI reasoning and human thinking. These findings are very valuable for AI educators and developers. Educators can use ChatGPT for quick Solution summaries and DeepSeek for detailed verification tailored to the student's profile. For developers, incorporating verification tools and maintaining consistency in notation and terminology are crucial for making AI a tool for deep, inclusive mathematical thinking rather than just providing answers.

5. REFERENCES

- Anua, S. M., Chong, J., Nasir, J. N. M., Zamri, M. M. M., Lihat, N., & Gulpi, N. A. (2025). Isu Keselamatan dan Etika Penggunaan Kecerdasan Buatan. *Journal of Energy and Safety Technology (JEST)*. <https://doi.org/10.11113/jest.v8i1.195>
- Arifin, M., Zulkarnain, I., & Ansori, H. (2025). The influence of artificial intelligence on critical thinking ability in mathematics: A systematic literature review. *Indonesian Journal of Science and Mathematics Education*. <https://doi.org/10.24042/ijsme.v8i1.24352>
- Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M., & Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior*, 148, 107903. <https://doi.org/10.1016/j.chb.2023.107903>
- Canonigo, A. M. (2024). Levering AI to enhance students' conceptual understanding and confidence in mathematics. *Journal of Computer Assisted Learning*, 40(6), 3215–3229. <https://doi.org/10.1111/jcal.13065>
- Copur-Gencturk, Y., Li, J., & Atabas, S. (2024). Improving Teaching at Scale: Can AI Be Incorporated Into Professional Development to Create Interactive, Personalized Learning for Teachers? *American Educational Research Journal*, 61(4), 767–802. <https://doi.org/10.3102/00028312241248514>
- Hubert, T., Mehta, R. S., Sartran, L., Horváth, M. Z., Žužić, G., Wieser, E., Huang, A., Schrittwieser, J., Schroecker, Y., Masoom, H., Bertolli, O., Zahavy, T., Mandhane, A., Yung, J., Beloshapka, I., Ibarz, B., Veeriah, V., Yu, L., Nash, O., . . . Silver, D. (2025). Olympiad-level formal mathematical reasoning with reinforcement learning. *Nature*, 651, 607 - 613. <https://doi.org/10.1038/s41586-025-09833-y>
- Hwang, S. (2022). Examining the Effects of Artificial Intelligence on Elementary Students' Mathematics Achievement: A Meta-Analysis. *Sustainability*, 14(20), 13185. <https://doi.org/10.3390/su142013185>
- Islam, K., Saidah, N. L., Yahya, S., & Syaikhuddin, M. M. (2026). Evaluating AI Tutor Interaction with Ambiguous Junior High Mathematics Questions Using Black-Box. *RIGGS: Journal of Artificial Intelligence and Digital Business*. <https://doi.org/10.31004/riggs.v5i1.6363>
- Macina, J., Daheim, N., Hakimi, I., Kapur, M., Gurevych, I., & Sachan, M. (2025). MathTutorBench: A Benchmark for Measuring Open-ended Pedagogical Capabilities of LLM Tutors. *ArXiv*, abs/2502.18940. <https://doi.org/10.48550/arxiv.2502.18940>
- Maulida, L., Nurossobah, P., Aura, B. A., Nengsih, E. D., & Rasilah, R. (2024). Improving The Effectiveness of Mathematics Learning Through Artificial Intelligence: Literature Review. *Journal of General Education and Humanities*. <https://doi.org/10.58421/gehu.v3i4.267>
- Moon, I. H. (2025). Benchmarking Large Language Models for Calculus Problem-Solving: A Comparative Analysis. *ArXiv*, abs/2504.13187. <https://doi.org/10.48550/arxiv.2504.13187>
- Opesemowo, O. A. G., & Ndlovu, M. (2024). Artificial intelligence in mathematics education: The good, the bad, and the ugly. *Journal of Pedagogical Research*. <https://doi.org/10.33902/JPR.202426428>
- Rahmatika, W., Mursalin, M., Saputra, E., & Fonna, M. (2025). The Influence of the Use of AI in Mathematics Learning at the High School Level: A Systematics Literature Review. *Electronic Journal of Education, Social Economics and Technology*. <https://doi.org/10.33122/ejeset.v6i2.1068>
- Sa'adah, N., Faizah, S., Sa'dijah, C., Khabibah, S., & Kurniati, D. (2023). Students' Mathematical Thinking Process in Algebraic Verification Based on Crystalline Concept. *Mathematics Teaching Research Journal*, 15(1), 90-107.

- Scarlatos, A., & Lan, A. S. (2023). Tree-Based Representation and Generation of Natural and Mathematical Language. 3714-3730. <https://doi.org/10.48550/arxiv.2302.07974>
- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7). <https://doi.org/10.29333/ejmste/13272>
- Xia, S., Li, X., Liu, Y., Wu, T., & Liu, P. (2024). Evaluating Mathematical Reasoning Beyond Accuracy. 27723-27730. <https://doi.org/10.48550/arxiv.2404.05692>
- Zhang, J., & Zhang, Z. (2024). AI in teacher education: Unlocking new dimensions in teaching support, inclusive learning, and digital literacy. *Journal of Computer Assisted Learning*, 40(4), 1871–1885. <https://doi.org/10.1111/jcal.12988>