TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH
Tập 23, Số 3 (2026): 682-694
Website: https://journal.hcmue.edu.vn
ISSN: 2734-9918

HO CHI MINH CITY UNIVERSITY OF EDUCATION
JOURNAL OF SCIENCE
Vol. 23, No. 3 (2026): 682-694
https://doi.org/10.54607/hcmue.js.23.3.4657(2026)

Research Article

ASSESSMENT OF CHATGPT'S MATH SOLVING ABILITY: THE CASE OF EXERCISES IN GRADE 12 MATH TEXTBOOKS IN VIETNAM

***Tang Minh Dung*[1*], *Le Hoang Quan*[2], *Nguyen Dinh Anh Hao*[1], *Nguyen Minh Ha*[1], *Le Ngoc Anh Tho*[1], *Pham Minh Thuy*[1]**

[1]*Ho Chi Minh City University of Education, Vietnam*

[2]*Dong Nai University, Vietnam*

[*]*Corresponding author: Tang Minh Dung – Email: dungtm@hcmue.edu.vn*

Received: December 24, 2024; Revised: April 17, 2025; Accepted: June 28, 2025

ABSTRACT

Artificial intelligence (AI) is transforming how we do things. In school, teachers use it to improve lessons or help students. Many math teachers pay attention to the ChatGPT version o1-mini since it can solve math problems well. This study examines how accurately ChatGPT o1-mini can solve math problems for 12th-grade students in Vietnam. We looked at 675 math problems in three 12th-grade math textbooks in Vietnam (Canh dieu, Chan troi sang tao, and Ket noi tri thuc voi cuoc song). The results showed that ChatGPT o1-mini could solve 53.9% of the exercises, and it did the best in Statistics. However, 16% of the responses provided by ChatGPT o1-mini had mistakes, and 30.1% of the responses required further explanation as ChatGPT o1-mini used methods or terms that did not belong in the curriculum. The study also found out which contents have mistakes or wrong ideas, in the answers given by ChatGPT o1-mini. These results give math teachers a clear view of how well AI performs and how correct its answers are. This will help them figure out how to use AI, specifically ChatGPT o1-mini, in their teaching.

Keywords: Artificial intelligence; ChatGPT; Generative AI; Math exercises

1. Introduction

Artificial Intelligence (AI) shows enormous potential in many Southeast Asian fields, with pronounced impacts on finance, healthcare, high technology, telecommunications, and education (Relmasira et al., 2023). AI is an important educational force in an era of rapid technological advancement. Artificial intelligence's application in education sets the scene for new teaching and learning methods in Mathematics and changes the old ones, as it offers previously impossible approaches (Yeo et al., 2024). Artificial intelligence technology enriches Mathematics education by providing innovative tools, resources, and methods that

Cite this article as: Tang, M. D., Le, H. Q., Nguyen, D. A. H., Nguyen, M. H., Le, N. A. T., & Pham, M. T. (2026). Assessment of ChatGPT's Math solving ability: The case of exercises in Grade 12 Math textbooks in Vietnam. *Ho Chi Minh City University of Education Journal of Science, 23*(3), 682-694. https://doi.org/10.54607/hcmue.js.23.3.4657(2026)

complement traditional teaching methods (Egara & Mosimege, 2024). While traditional Mathematics teaching methods, often including textbooks and teachers' lectures, have not fulfilled the growing needs of learners yet, AI brings a natural sense of enjoyment and entertainment to the learning process, which allows students to analyze the content further and receive responses. The potential benefits of using AI in education are significant. One of the most enormous benefits of AI in education is personalized learning. It can lead to better academic outcomes, as students can learn at their own pace and by their own learning styles (Harry, 2023). AI technologies can be used to analyze students' learning processes, including interactive content, learning behaviors, test results, and perceptions of learning, provide instant support or feedback to each student, as well as make recommendations to teachers on how to improve teaching content and strategies (Hwang & Tu, 2021).

Although there had been significant attention to artificial intelligence previously, it was not until the emergence of ChatGPT – a form of artificial intelligence based on OpenAI's GPT (Generative Pretrained Transformer) model – in 2022 that discussions regarding the application of AI in education and academia indeed surged. This surge was primarily due to ChatGPT's ability to generate written content closely resembling human text (Wardat et al., 2023). With an extensive repository of resources and contextual understanding capabilities, ChatGPT is highly versatile for content creation, thereby supporting the execution of educational tasks (Javaid et al., 2023). According to the 2018 Mathematics curriculum for general education, the application of advancements in cutting-edge digital technology in teaching is of interest, contributing to developing essential mathematical competencies for students. Specifically, the Ministry of Education and Training (2018) has identified the ability to use mathematical tools and learning aids as one of five key mathematical competencies for students to be developed by teachers.

Numerous studies in Vietnam and worldwide have shown that ChatGPT is advantageous in education. Specifically, ChatGPT can do Literature and Math tests (Le et al., 2023). However, the scores are not as good as when students do the tests themselves. ChatGPT can also assist in the teaching of Physics in Vietnamese high schools, as mentioned in Liang et al. (2023). It can give solutions, explain ideas, and make learning attractive for students. It helps teachers do their job better. Another study, Niloy et al. (2024), tried to figure out what makes students use ChatGPT. The findings demonstrated the associations between the intention of the students to use ChatGPT and the actual engagement with the platform. Although some people still worry about the accuracy of ChatGPT, it remains a helpful source for students by suggesting things like syntax and grammar or generating practice exercises and quizzes for math, physics, language, or literature. ChatGPT also explains problems step-by-step and thus, it promotes problem-solving skills and thinking analytically and creatively (Rahman & Watanobe, 2023).

In September 2024, the ChatGPT o1-mini version was launched by OpenAI. This was a breakthrough in text data processing since this version is designed to spend more time thinking before responding, and thus, it can reason through complex tasks and solve harder problems than the previous version. When it comes to education, notably in math education, ChatGPT o1-mini is going to change a lot of things. It will help us work with text data in ways that make

teaching and learning better, such as handling lesson notes, student answers, and even feedback. However, so far, no studies have evaluated the success rate of ChatGPT o1-mini when solving exercises in Grade 12 Math textbooks in Vietnam. This raises the research question: To what extent can ChatGPT o1-mini provide accurate solutions to the exercises in the current Grade 12 Math textbooks? What is the scope of ChatGPT o1-mini's effective use?

Evaluating the ability of ChatGPT o1-mini when solving 12th-grade Math exercises not only helps determine the reliability of this technology but also contributes to providing awareness when using AI in an educational environment. If ChatGPT o1-mini proves to be highly effective, it could be a useful tool for teachers and students. Teachers could use it to assist in lesson planning, while students might use it to aid their learning. On the other hand, if ChatGPT o1-mini is not accurate or does not fit with standard programs, changes and additions are necessary. This study aims to give recommendations on how teachers and students can better understand and use ChatGPT o1-mini for teaching and learning Math in high schools, in Vietnam. The goal is to help teachers and students use ChatGPT o1-mini effectively for Math.

2. Research Methodology

Stage 1: Identifying Exercises to Survey

The study looked at 675 exercises from three current Grade 12 textbooks. These exercises include mathematics content domains such as Calculus, Geometry and Measurement, Statistics, and Probability. We only used problems that are based on text and excluded those that have pictures, like diagrams or graphs. The reason for this is that ChatGPT cannot handle images well in its current version. This means it might get the idea or miss important parts of problems with pictures.

Stage 2: Constructing the prompt structure

We designed a sample prompt to obtain responses from ChatGPT o1-mini for solving math problems (Figure 1). This prompt comprises two main components. First, ChatGPT is instructed to "act" as a Mathematics teacher with extensive knowledge of the 2018 Vietnamese General Education Curriculum for Mathematics. Second, ChatGPT will solve the provided exercise problems and return the results as LaTeX source code.

You are a mathematics teacher teaching Grade 12 at an exceptionally outstanding professional level and are highly proficient in LaTeX source code. You also have an in-depth understanding of how to solve mathematics exercises according to the "2018 Vietnamese General Education Mathematics Curriculum." Please help me solve the following math problem using the latest solution methods. Always present only the problem and the solution without omitting any part, ensuring that the problem and the solution are displayed when compiling the LaTeX source code. Always place the problem and solution within the `document` environment. Note that when dealing with mathematical content, always write it professionally using LaTeX source code. Do not display it as formulas. When you want to list items, always use the `itemize` environment for the listed content and the related components in the problem and solution. The list of questions answered in the solution section must match the questions in the problem section. At the same time, the solutions to the ideas in the problem must always be placed together in the `itemized` environment corresponding to the list of answered questions in the solution section. Always use the following packages to utilize all necessary symbols and functionalities: amsmath, amssymb, amsthm, mathtools, graphicx, hyperref, tikz, array,... Mathematical symbols, such as integrals, sums, limits, etc., must always be presented accurately and correctly. Always use environments like align, align* to align lines of formulas when alignment is desired instead of using other environments. Additionally, use the tabular environment to create tables. Always use the \displaystyle command for formulas that need emphasis or to appear independently. Always use the \label and \ref commands to reference formulas, facilitating easier problem tracking efficiently. Always use parentheses or environments like cases when handling multiple different cases. Furthermore, since you are writing in Vietnamese, please add \usepackage[utf8]{vietnam} in the LaTeX source code. Do not add a title, author, date, or make the title.

"The problem is entered here."

Figure 1. *A sample prompt to request ChatGPT to solve math problems*

Stage 3: Loading the prompt into ChatGPT

We proceeded to load the prompt segments according to the aforementioned format, adding the problem statements to the ChatGPT o1-mini version to initiate the exercise-solving process. After ChatGPT processed each prompt, all generated LaTeX codes, including the problem statements and solutions, were stored in the database via a form supported by Google Forms. Each textbook organized this data to facilitate management and subsequent retrieval.

Stage 4: Classifying the results obtained from ChatGPT

In this stage, we accessed and reviewed the data using Overleaf. Subsequently, the process of categorizing and labeling all the responses provided by ChatGPT was done manually. The dataset was divided into four sections according to the order numbers in the Excel file exported from Google Forms. A team of four members was assigned to cross-check two sections per person. The results were labeled into three categories: (1) responses with errors, (2) correct responses that align with the current national curriculum, and (3) correct responses that may confuse students. In case (3), the confusing elements of the knowledge were explicitly identified. In case (1), errors were documented for qualitative analysis in Stage 6.

Stage 5: Descriptive statistics of the number of each type of response

The research team conducted a statistical analysis of the data compiled and categorized in Stage 4. We counted the number of responses in each category and visualized the results using pie charts to identify trends based on the proportion of each response type.

Stage 6: Thematic analysis

Based on the qualitative data obtained from cases (1) and (3), we organized the exercise problems, confusing concepts, and common errors into a cross-tabulation table. This approach allowed us to conclude ChatGPT o1-mini's problem-solving capabilities.

3. Result and discussion

3.1. Statistics of ChatGPT responses o1-mini

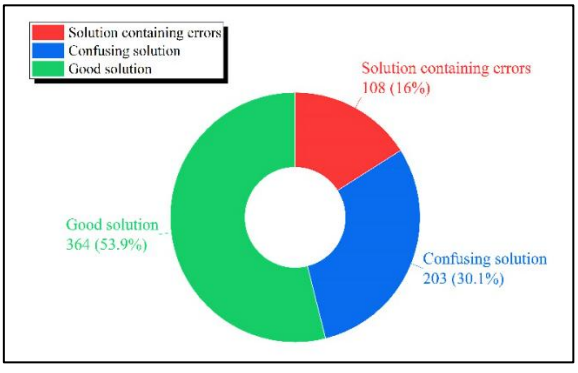

Figure 2. *Distribution of the quality of ChatGPT o1-mini's solutions for problems in Grade 12 Mathematics Textbooks in Vietnam*

The figure presents ChatGPT o1-mini's performance in solving problems from these textbooks, categorized into three groups. First, "a solution containing errors" (108 problems) – a solution in which mistakes in reasoning or execution lead to an incorrect or incomplete

outcome. The example below showcases a notable error: ChatGPT o1-mini did not remove *x=-1* (where the denominator is 0) from the domain, leading to an incorrect outcome.

$y' = \dfrac{x^2+2x-2}{(x+1)^2}$; $y' = 0 \Rightarrow x = -1 \pm \sqrt{3}$. *Kết luận:* y *đồng biến trên* $\left(-\infty, -1-\sqrt{3}\right)$, $\left(-1+\sqrt{3}, +\infty\right)$ *và* y *nghịch biến trên* $\left(-1-\sqrt{3}, -1+\sqrt{3}\right)$.

Next, "a confusing solution" (203 problems). The second group comprises solutions that employ unfamiliar knowledge (i.e., knowledge not covered in the 2018 Vietnamese Mathematics curriculum) or explanations likely to perplex students, which is illustrated by the following example: "Linearly independent" and "basis vector" are unfamiliar concepts for students in this case.

> *Ba vector* $\vec{a}, \vec{b}, \vec{c}$ *khác* $\vec{0}$ *và vuông góc từng đôi nên chúng độc lập tuyến tính (linearly independent). Từ đó, ba vector* $\vec{a}, \vec{b}, \vec{c}$ *có thể được sử dụng làm vector cơ sở (basis vector) của không gian* $Oxyz$ *, nên ta lập được hệ trục* Ox, Oy, Oz *song song với các vector này.*

Finally, "a good solution" (364 problems). The third group comprises solutions that are accurate, clear, and well-structured in alignment with the 2018 Vietnamese Mathematics curriculum, as exemplified by the excerpt below:

$\int 2e^x \mathrm{d}x = 2\int e^x \mathrm{d}x = 2e^x + C$, trong đó C là hằng số tùy ý.

This study looked at a lot of problems, so the results are really good at showing what ChatGPT o1-mini can do. The results give us a fairer view of how well ChatGPT o1-mini does with high school mathematics problems. ChatGPT o1-mini is what we are talking about here. This study helps us understand how it performs with these kinds of problems.

The figure shows that "a good solution" accounts for about 53.9% of the results. This shows that ChatGPT o1-mini is good at giving precise answers. These results are similar to what other people have found out, that ChatGPT can handle basic high school math problems (Vu, 2024) and answer questions that need recognition and comprehension (Le et al., 2023) pretty well.

This shows that ChatGPT o1-mini still has trouble with some math problems. These mistakes are mainly caused by the failure to understand what needs to be done, failure to solve problems, or sometimes leaving out important information. Each problem needs to be looked at to see what is going wrong. We did not just want to check if the answers were correct or not. We also wanted to see if the solutions made sense and were easy to understand. If they were suitable for high school students. We found out that one-third of the solutions to 203 problems were hard to understand. This is because ChatGPT o1-mini does not know who it is talking to, and it often uses explanations that are too hard for high school students. Another reason is that the o1-mini version tries to give answers and use general methods to solve problems quickly. This can make it hard for students to understand, especially if it does not explain things clearly or show the steps it took to get the answer.

Table 1. *Analysis of factors causing errors in ChatGPT o1-mini's solutions by mathematical problem types*

		Common Error Factors			
		Errors in applying mathematical methods and techniques	**Errors in mathematical reasoning and proof**	**Errors in vector representation operations**	**Errors in mathematical computations and operations**
Problems	Algebra and Some Elements of Calculus	59	6	0	16
	Geometry and Measurement	2	15	2	8
	Statistics and Probability	0	0	0	2

3.2. Errors in ChatGPT o1-mini's solution

The table shows that the highest concentration of errors occurs in Algebra and Some Elements of Calculus, with 59 problems involving errors related to mathematical methods and techniques. For example, when addressing problems involving the determination of vertical asymptotes of a function, ChatGPT frequently identifies the roots of the denominator and assumes them to be vertical asymptotes rather than employing limit-based analysis to verify their existence. Additionally, there are 16 errors related to mathematical calculations and operations, such as errors in addition and subtraction:

$$... = (-9 + 4.5 + 18) - \left(-\frac{8}{3} + 2 - 10 \right) = 13.5 - (-1.333...) .$$

Six mistakes in mathematical reasoning and proof are found. For instance, it is a statement without an explanation, as in the following case:

$F'(x) = e^{\sin x} \cdot \cos x \neq e^{\cos x} = f(x)$ với mọi $x \in \mathbb{R}$.

Errors associated with methods and mathematical techniques primarily involve determining asymptotes of functions without employing limited operations, and the incomplete execution of procedures when determining a given function's local maxima/minima or the maximum/minimum values. This observation aligns with the assertion that AI ChatGPT has not provided detailed solutions for problems involving graphs and integrals (Vu, 2024). Conversely, this study contradicts the view that ChatGPT performs relatively well on questions at the recognition and comprehension levels related to the content strands of the greatest and least values of functions and asymptotes (Le et al., 2023). This difference may reflect updates to ChatGPT over time, which seem to have influenced the way it handles and solves mathematical problems.

The most prevalent error in the Geometry and Measurement module involves mathematical reasoning and proof (15 problems). For instance, ChatGPT argues that for a

line to be perpendicular to a plane, the direction vector of the line must be perpendicular to the normal vector of the plane rather than parallel to it. Secondly, errors in mathematical computations and operations are found in eight problems, such as the wrong formula in determining the angle between a line and a plane:

$$\cos\theta = \frac{|\vec{u}\cdot\vec{n}|}{\|\vec{u}\|\cdot\|\vec{n}\|}.$$

Next, there are two errors involving mathematical methods and techniques, such as the wrong general equation of a plane:

$\vec{n}\cdot(\vec{r}-\vec{r_0}) = 0$, trong đó $\vec{r_0}$ là vị trí điểm thuộc mặt phẳng.

Finally, there are two problems containing errors in vector representation operations, such as:

$\overrightarrow{CA'} + \overrightarrow{AD} = -\overrightarrow{AC} + \overrightarrow{AD}$.

This finding is consistent with the conclusion that ChatGPT cannot systematically present solution steps and often makes random errors across different solution attempts (Le et al., 2023).

In the Statistics and Probability module, the number of errors is minimal, with only two problems involving errors related to mathematical computations and operations, and no other types of errors were observed. For example, ChatGPT provided the correct formula for calculating the average value, but still made mistakes when substituting numerical values and obtaining the final result. Earlier studies by Le et al. (2023), Vu (2024), and Wardat et al. (2023) focused mainly on ChatGPT's performance in Algebra, Some Elements of Calculus, Geometry, and Measurement. The study is more comprehensive since we analyzed all mathematics context domains in the established curriculum.

The mistakes may come from how ChatGPT gives answers. It just tries to answer without going through each step carefully, leading to the appearance of mistakes when something needs thinking, special techniques, or many steps of calculation. Another problem is with geometry problems. ChatGPT does not actually draw or check figures; it only relies on written descriptions. This may lead to confusion about symbols or how things are related in space and bring about the wrong depiction of geometric objects. ChatGPT o1-mini seems quite proficient at Statistics and Probability problems when it just has a few mistakes. It may be that we have excluded problems having pictures, like diagrams or graphs, which might have decreased the number of documented errors.

3.3. Common Confusing Factors in ChatGPT's o1-mini Solution

ChatGPT sometimes used complicated words or ways to explain things. For example, it is able to solve Grade 9 math problems using ideas such as derivatives, which are way beyond that level (Le et al., 2023). The o1-mini version had similar issues, too. Some answers were above Grade 12 or points that could be confusing to the students. Table 2 summarizes these cases and shows the main sources of confusion found in ChatGPT o1-mini's solutions across three mathematical topics. These factors are categorized into three main categories: Mathematical Notation and Representation, Mathematical Methods and Formulas, and Terms and Concepts.

Table 2. *Analysis of confusing factors in ChatGPT o1-mini's solutions for mathematical problem types*

		Common Confusing Factors		
		Mathematical Notation and Representation	**Mathematical Methods and Formulas**	**Terms and Concepts**
		Count	**Count**	**Count**
Problems	Algebra and Some Elements of Calculus	83	16	53
	Geometry and Measurement	34	65	5
	Statistics and Probability	0	0	0

Among these areas, Algebra and Some Elements of Calculus have the highest number of confusing factors, accounting for 152 problems containing elements that could lead to confusion. This is followed by Geometry and Measurement, which includes 104 such problems. Notably, no confusing factors were reported for problems related to Statistics and Probability, indicating that ChatGPT o1-mini provided solutions that closely aligned with the 12th-grade curriculum for this topic.

In Mathematical Notation and Representation, the most confusing problems are in Algebra and Some Elements of Calculus, with 83 problems. As demonstrated in the following solution, the most commonly confusing notation is the derivative notation $\mathrm{d}y/\mathrm{d}x$, which is no longer part of the 2018 curriculum.

$$y = -x^3 + 3x^2 - 9x + 1 \Rightarrow y' = \frac{\mathrm{d}y}{\mathrm{d}x} = -3x^2 + 6x - 9 .$$

Geometry and Measurement recorded 34 problems in this category, with frequent confusion arising from representations such as vector coordinates in column matrix form and using the norm symbol for vector magnitudes, as illustrated in the following solution:

$$\left\|\overrightarrow{u_2}\right\| = \sqrt{1^2 + 1^2 + \left(-\sqrt{2}\right)^2} = \sqrt{1+1+2} = \sqrt{4} = 2 .$$

For the Mathematical Methods and Formulas category, Geometry and Measurement have the highest number of problems involving confusing factors, with 65 problems. ChatGPT o1- mini's solutions often use several advanced methods and formulas in geometry, such as matrix determinants for calculating the cross product and formulas for calculating vector coordinates or distances between parallel planes, which are beyond the scope of the typical high school curriculum, as shown in the following solution:

Tính tích có hướng: $\vec{n}_P = \begin{vmatrix} i & j & k \\ 3 & 2 & -1 \\ 1 & 1 & -1 \end{vmatrix}$.

On the other hand, the Algebra and Some Elements of Calculus area recorded 16 problems. A typical example is ChatGPT o1-mini, frequently using second-order derivatives to identify minima and maxima, which are no longer included in the 2018 Mathematics curriculum:

Tại $x = 2 : y''(2) = 12(2) + 6 = 30 > 0 \Rightarrow x = 2$ là điểm cực tiểu.

Algebra and Some Elements of Calculus recorded 53 problems in the Terms and Concepts category. For instance, the following solution includes advanced terms, which may be unfamiliar to students who have not yet encountered them in their curriculum:

Áp dụng định lí Newton-Leibniz (Định lí cơ bản của Giải tích), ta có: $\int_1^3 f'(x)dx = f(3) - f(1)$.

Geometry and Measurement recorded 5 problems involving confusing terms or concepts such as linear independence, the basis of a vector space, and orthogonal basis, which are fundamental to more advanced geometry and linear algebra topics. However, they are not part of the high school curriculum. Their use can be seen in the following solution:

Ba vectơ đôi một vuông góc và đều khác vectơ không nên suy ra chúng là độc lập tuyến tính.

The accuracy and relevance of ChatGPT's responses significantly depend on the clarity and specificity of the prompts provided (Le et al., 2023). However, our research reveals that even when the prompt explicitly defines the role and scope, "You are a mathematics teacher teaching Grade 12 with an exceptionally outstanding professional level, and you are highly proficient in LaTeX source code. You also have an in-depth understanding of how to solve mathematics exercises according to the 2018 Vietnamese General Education Mathematics Curriculum. ChatGPT o1-mini's responses still include confusing elements, such as content that exceeds or deviates from the specified curriculum. This issue may be blamed on the model's training data, which includes mathematical content beyond the high school curriculum. As a result, ChatGPT o1-mini sometimes gives detailed solutions that rely on methods or formulas beyond the level expected of high school students.

4. Conclusion

In this study, we examine how accurately ChatGPT o1-mini can solve math problems for 12th-grade students in Vietnam. The primary goal is not only to assess the reliability of this technology but also to provide insights on using AI in the educational environment and offer specific recommendations to enhance teachers' and students' understanding of applying AI technology, particularly ChatGPT o1-mini, in teaching and learning Mathematics in high schools in Vietnam. These recommendations aim to help teachers optimize their teaching processes. At the same time, students can leverage AI as a practical learning tool, ultimately improving the quality of teaching and learning Mathematics in the context of digital transformation today.

We analyzed the results and found that ChatGPT's ability to solve the provided math problems correctly was over 53.9%. Notably, the ability to solve probability and statistics-related problems was flawless for the o1-mini version. These findings do not fully align with the results of the author group Guler, who claimed that ChatGPT 3.5 and 4 made significant errors in probability-related questions (Guler et al., 2024). The study confirms the progress made by ChatGPT, thanks to reinforcement learning mechanisms, which allow the evaluation of multiple responses to select the most appropriate one (Dwivedi et al., 2023). However, like any technology, it still has limitations, which result in inaccurate outputs (Wardat et al., 2023). The study's findings indicate that ChatGPT's responses can sometimes

contain errors and may only sometimes provide the most optimal or practical solutions. This aligns with another study showing that some participants expressed concerns about the accuracy and reliability of ChatGPT, as it could provide incorrect or incomplete solutions to mathematical problems (Wang et al., 2023).

This study has several acknowledged limitations. Initially, the mathematics exercises selected for assessment represent only a subset of the Vietnamese Grade 12 curriculum. The implication, then, is that the research does not cover ChatGPT's capacity to tackle a whole range of math topics, especially those that involve diagrams, graphs, and tables. This, in turn, limits the relevance of the results mainly to text-based problem formats. For another, it has been limited to just one AI app (the ChatGPT version o1-mini), but without comparing to other generative AI engines. In order to estimate the overall performance of AI in educational domains, future research needs to offer comparable results using complementary resources, such as Google Gemini or Microsoft Copilot. Third, this study was devoid of the user experience dimension, namely, responses from teachers and students to the AI solutions that were produced. It thus remains uncertain whether it is possible that users, especially students, who are used to conventional academic terms, may find themselves confused by terminology or a method that is outside the scope of their past classroom preparation. An appreciation for these in-practice problems is crucial not only to critically appraise the utility of AI-based learning in the context of real-life use but also to identify opportunities for further refining in terms of its pedagogical pertinence and classroom relevance.

Further research efforts on prompt design are needed to maximize the accuracy, context sensitivity, and context relevance of ChatGPT's responses. It should be emphasized that reaching the universal "optimal" prompt is virtually impossible, as the inputs from the user are so complex and dynamic. The best possible approach would be to try repeated iterations with a range of prompts. In this case, we employed an explicit prompt that defined the user role, the education context, and expected formatting in LaTeX. It has also been found that the type and pattern of the prompts influence output quality evidently (Shi et al., 2020). Therefore, the development of prompt engineering skills (e.g., role, goal identification, contextual boundaries, and iterative dialogues refining output) is a key aspect for the effective use of ChatGPT in an educational context (Ingley & Pack, 2023). Good prompts also support producing more sophisticated, relevant, and instructional information that is more accurate. Effective prompts can also lead to the generation of more nuanced, pertinent, and instructive information, ultimately resulting in greater accuracy. Still, ChatGPT has a few drawbacks here, and more advanced work may take us further, as the tasks grow more complex, particularly in terms of different formats, mathematical symbols, and symbolic representations. Potential areas that could be useful for future research are interviewing teachers or students regarding their experience with ChatGPT in the context of mathematics education, in-depth or through surveys. Previous research indicates that the interactive and conversational aspect of AI-driven educational tools like ChatGPT can add to the learning experience by enhancing motivation among students and thereby establishing

a more stimulating environment (Alshahrani, 2023; Rawas, 2024). In the context of teaching, some educators suggested that ChatGPT tools could be used to better deliver instruction, revealing the potential for technology to bolster teachers' pedagogical activity (Jeon & Lee, 2023). ChatGPT could support teachers in their roles both inside and outside of the classroom by assisting them in developing personalised learning experiences and making prompt responses. Future research should then endeavour to explore ways in which AI-generated content can be adapted to teachers' purposes and students' learning preferences. And based on the expected results, some practical suggestions to teachers of ChatGPT may include recommending them use the tool for generating test items or suggesting techniques for alternative solutions and, at the same time, to facilitate students to systematically analyze those AI-generated products under standard parameters to validate the reliability and further deepen the understanding of what the machine is doing.

Equipping people with AI and information technology knowledge is really important for helping them get the most out of future opportunities. AI can be helpful when it comes to solving problems and finding solutions. It will never be able to take the place of teachers. The thing is, AI just cannot tell how well students understand things, and a teacher can change the way they teach to fit each student. Even when you give AI clear instructions, it can still have trouble thinking and reasoning because of the different ways that people teach and learn math in different countries. There is no doubt that the use of AI in education will increase in the future. Therefore, it is crucial that we give educators the knowledge and skills they need to use AI in a way that benefits students and prevents potential issues. In this way, educators can use AI to improve the learning process and ensure that students benefit as much as possible. Consequently, this will guarantee the long-term advancement of the educational system overall and improve the quality of mathematics instruction.

❖ ***Conflict of Interest:*** *Authors have no conflict of interest to declare.*

REFERENCES

Alshahrani, A. (2023). The impact of ChatGPT on blended learning: Current trends and future research directions. *International Journal of Data and Network Science*, *7*(4), 2029–2040. https://doi.org/10.5267/j.ijdns.2023.6.010

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Egara, F. O., & Mosimege, M. (2024). Exploring the integration of artificial intelligence-based ChatGPT into Mathematics instruction: Perceptions, challenges, and implications for educators. *Education Sciences*, *14*(7), 742. https://doi.org/10.3390/educsci14070742

Guler, N. K., Dertli, Z. G., Boran, E., & Yildiz, B. (2024). An artificial intelligence application in mathematics education: Evaluating ChatGPT's academic achievement in a mathematics exam. *Pedagogical Research*, *9*(2), em0188. https://doi.org/10.29333/pr/14145

Harry, A. (2023). Role of AI in education. *Interdisciplinary Journal and Humanity (INJURITY)*, *2*(3), 260–268. https://doi.org/10.58631/injurity.v2i3.52

Hwang, G.-J., & Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, *9*(6), 584. https://doi.org/10.3390/math9060584

Ingley, S. J., & Pack, A. (2023). Leveraging AI tools to develop the writer rather than the writing. *Trends in Ecology & Evolution*, *38*(9), 785–787. https://doi.org/10.1016/j.tree.2023.05.007

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, *28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Le, A. V., Bui, T. D., Le, Q. Q., & Vu, V. L. (2023). The ability of ChatGPT to conduct periodic exams in Mathematics and Literature at the secondary school level: Research findings and some initial recommendations. *Vietnam Journal of Educational Sciences*, 19(02). https://doi.org/10.15625/2615-8957/12310201

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

Ministry of Education and Training. (2018). *Circular of Decree No. 32/2018/TT-BGDDT dated December 26, 2018 for promulgating General Education program*. Hanoi.

Niloy, A. C., Bari, M. A., Sultana, J., Chowdhury, R., Raisa, F. M., Islam, A., Mahmud, S., Jahan, I., Sarkar, M., Akter, S., Nishat, N., Afroz, M., Sen, A., Islam, T., Tareq, M. H., & Hossen, M. A. (2024). Why do students use ChatGPT? Answering through a triangulation approach. *Computers and Education: Artificial Intelligence*, *6*, 100208. https://doi.org/10.1016/j.caeai.2024.100208

Rahman, Md. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, *13*(9), 5783. https://doi.org/10.3390/app13095783

Rawas, S. (2024). ChatGPT: Empowering lifelong learning in the digital age of higher education. *Education and Information Technologies*, *29*(6), 6895–6908. https://doi.org/10.1007/s10639-023-12114-8

Relmasira, S. C., Lai, Y. C., & Donaldson, J. P. (2023). Fostering AI literacy in elementary Science, Technology, Engineering, Art, and Mathematics (STEAM) Education in the age of generative AI. *Sustainability*, *15*(18), 13595. https://doi.org/10.3390/su151813595

Shi, B., Huang, L., & Lu, X. (2020). Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing*, *37*(3), 361–388. https://doi.org/10.1177/0265532220911626

Vu, N. H. (2024). Applying artificial intelligence in high schools. *Vietnam Journal of Educational Sciences,* 20(S2). https://doi.org/10.15625/2615-8957/12420207

Wang, J., Liu, Z., Zhao, L., Wu, Z., Ma, C., Yu, S., Dai, H., Yang, Q., Liu, Y., Zhang, S., Shi, E., Pan, Y., Zhang, T., Zhu, D., Li, X., Jiang, X., Ge, B., Yuan, Y., Shen, D., … Zhang, S. (2023). Review of large vision models and visual prompt engineering. *Meta-Radiology*, *1*(3), 100047. https://doi.org/10.1016/j.metrad.2023.100047

Wardat, Y., Tashtoush, M. A., Ali, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2286. https://doi.org/10.29333/ejmste/13272

Yeo, S., Moon, J., & Kim, D.-J. (2024). Transforming mathematics education with AI: Innovations, implementations, and insights. *The Mathematical Education*, *63*(2), 387–392. https://doi.org/10.7468/MATHEDU.2024.63.2.387

ĐÁNH GIÁ NĂNG LỰC GIẢI TOÁN CỦA CHATGPT: TRƯỜNG HỢP CÁC BÀI TẬP TRONG CÁC SÁCH GIÁO KHOA TOÁN LỚP 12 Ở VIỆT NAM

Tăng Minh Dũng[1*], Lê Hoàng Quân[2], Nguyễn Đình Anh Hào[1], Nguyễn Minh Hà[1], Lê Ngọc Anh Thơ[1], Phạm Minh Thuỳ[1]

[1]Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam

[2]Trường Đại học Đồng Nai, Việt Nam

[*]Tác giả liên hệ: Tăng Minh Dũng – Email: dungtm@hcmue.edu.vn

Ngày nhận bài: 24-12-2024; Ngày nhận bài sửa: 17-4-2025; Ngày duyệt đăng: 28-6-2025

TÓM TẮT

Trí tuệ nhân tạo (AI) đang thay đổi cách chúng ta làm việc. Ở trường học, giáo viên sử dụng nó để cải thiện tiết học cũng như hỗ trợ học sinh. Nhiều giáo viên toán dành sự chú ý cho ChatGPT phiên bản o1-mini vì khả năng giải toán tốt. Nghiên cứu này xem xét độ chính xác của ChatGPT o1-mini khi giải toán lớp 12 tại Việt Nam. Chúng tôi phân tích 675 bài toán từ ba bộ sách giáo khoa toán lớp 12 của Việt Nam (Cánh diều, Chân trời sáng tạo, Kết nối tri thức với cuộc sống). Kết quả cho thấy ChatGPT o1-mini giải đúng 53,9% số bài, với kết quả tốt nhất ở phần Thống kê. Tuy nhiên, khoảng 16% lời giải có sai sót và 30,1% câu trả lời cần làm rõ thêm vì sử dụng phương pháp hoặc thuật ngữ ngoài chương trình. Nghiên cứu cũng chỉ ra các chủ đề thường xảy ra lỗi sai và hiểu nhầm trong câu trả lời của ChatGPT o1-mini. Những phát hiện này giúp giáo viên toán có cái nhìn khách quan về hiệu năng và độ chính xác của trí tuệ nhân tạo, từ đó sử dụng ChatGPT o1-mini và các công cụ trí tuệ nhân tạo khác một cách hiệu quả và phù hợp trong giảng dạy.

Từ khóa: trí tuệ nhân tạo; ChatGPT; AI tạo sinh; bài tập toán