

## **From problem solver to teaching partner: Rethinking the role of AI in high school mathematics competition education**

Arya Fu, Chin Khang Lee

*Shanghai Experimental Foreign Language School*

This study aims to evaluate the potential of ChatGPT-4o in mathematics competition teaching. By conducting a preliminary test of AI's analytical ability, personalized guidance ability, and learning plan generation ability, we explore its feasibility and limitations in teacher preparation, student self-study, and competition preparation. Experimental methods and text analysis methods were adopted to analyze ChatGPT-4o's answers to American mathematics competition questions under different prompts. It was found that ChatGPT-4o had a low accuracy rate when solving problems without prompts and lacked effectiveness in solving geometry problems, but it also showed the ability to provide effective personalized guidance in problem classification and knowledge point summary.

**Keywords: Generative AI; mathematics competition; personalized learning**

### **Introduction**

As an important part of mathematics education, mathematics competitions are designed to stimulate students' interest in mathematics learning in the form of extracurricular activities, and provide a platform for students of different ages to develop logical thinking and problem-solving skills, which has far-reaching significance for students' academic development and future career planning (Inglis & Attridge, 2020).

From basic calculation and problem-solving skills to advanced proof techniques, mathematics competitions assess students' mathematical abilities and provide students with stimulating and challenging problems beyond the regular curriculum. In some education systems, students' performance and achievements in math competitions are also considered important advantages for university admissions and scholarship applications.

From the perspective of actual teaching needs, students preparing for competitions have different mathematical learning foundations and experiences. Due to time and resource constraints, it is difficult for a unified school curriculum to meet these personalized needs (Radović et al., 2019). Under the global wave of educational technology, artificial intelligence (AI) provides a new path for personalized learning. Although AI has demonstrated its ability to solve complex mathematical problems, its teaching potential in high school mathematics competitions has received little attention (Hwang & Tu, 2021). This study aims to rethink the role of AI—from problem solver to teaching partner—by examining the application of AI in supporting personalized teaching in the American Mathematics Competition (AMC).

## Literature Review

Since traditional classroom teaching is difficult to meet the individual needs of each student, technology empowerment has become a key direction for personalized learning. The U.S. Department of Education (2017) pointed out that educational technology can reduce the burden on teachers and provide personalized learning paths. Scholars such as Insorio & Librada (2020) and Karakozov & Ryzhova (2019) also believe that the implementation of modern school education programs, including mathematics education, needs to consider the principles of personalized learning, actively use digital technology, and interactive means to ensure a high level of mathematics education.

Studies have shown that AI technology can improve students' academic performance and higher-order thinking skills (Voskoglou & Salem, 2020) and be used to diagnose students' learning problems and provide personalized feedback to help teachers optimize teaching design (Chen et al., 2020; Bywater et al., 2019; Hung et al., 2014). However, existing research on the application of AI in mathematics education mainly focuses on general mathematics courses, with less research on competition courses and research on high school students and teachers is relatively limited (Hwang & Tu, 2021).

Students of different grades have different needs in the same course and need personalized guidance from teachers. AI is a technical tool that can be used to support teachers in tutoring mathematics competitions. Existing research shows that AI can help teachers generate competition questions with difficulty similar to American Invitational Mathematics Examination (Zhao et al., 2025), which can help teachers prepare competition question banks based on student level needs and reduce the amount of preparation for competition teachers.

The purpose of this study is to fill the research gap in the use of artificial intelligence in personalized learning for high school mathematics competitions, explore how AI can optimize competition preparation strategies, improve students' mathematical thinking abilities, and provide new directions for the future development of intelligent educational technology. Three research questions are put forward:

RQ1: How accurate is AI in analyzing mathematics competition problems?

RQ2: How effective is AI in offering personalized problem-solving guidance?

RQ3: Can AI generate tailored learning plans based on student performance?

## Theoretical Framework

To explore the application of generative AI in preparation for the Mathematics Competition, this study adopted the AI-TPACK framework, which integrates technology, teaching, and content knowledge (Mishra, 2023). TPACK (Technological Pedagogical Content Knowledge) is a teacher professional development framework that emphasizes the dynamic integration and synergy between technology, teaching methods, and subject content, rather than simple knowledge superposition (Koehler & Mishra, 2009). The AI-TPACK framework responds to the research gap in the application of AI in personalized learning for high school mathematics competitions mentioned in the literature review (Hwang & Tu, 2021), and provides theoretical guidance for how teachers can design personalized teaching materials and improve students' competition performance and learning engagement. The AI-TPACK framework takes generative AI as its core and expands the traditional TPACK model

(Mishra & Koehler, 2006). Based on the actual situation of this research case, it is reflected in the following three dimensions:

- Technical knowledge (TK): Mathematics teachers' understanding of various generative artificial intelligence knowledge and their ability and proficiency in using AI tools.
- Pedagogical knowledge (PK): The teaching methods and approaches used by teachers in lesson preparation and teaching are based on students' cognitive abilities and learning situations, such as topic division, tiered exercises, instant feedback, and personalized guidance.
- Content knowledge (CK): Teachers should not only understand the core mathematical content of Mathematics Competition, but also be familiar with the syllabus of high school mathematics in order to understand the learning progress of students in different periods and grades. They should also be familiar with the examination styles of different mathematics competitions.

## Research Design

This study is a pilot study for subsequent studies. The AMC problem was selected as the research object mainly based on its representativeness in the international mathematics competition system. International competitions such as Math Kangaroo focus on the fun of mathematics and the cultivation of thinking, while some regional competitions are organized according to the International Mathematical Olympiad (IMO) syllabus to select high-level players.

The study is based on an international high school in Shanghai that adopts the A Level and AP curriculum system. Students are encouraged to participate in various international mathematics competitions, including the American Mathematics Competition (AMC), the United Kingdom Mathematics Trust - Senior Mathematical Challenge (SMC), Math Kangaroo (MK), Euclid Mathematics Competition, and other extracurricular activities they believe are interesting and valuable.

In these competitions, since the AMC competition is scheduled in October every year, students often need to start preparing for the exams in the summer, and there is a lack of teacher guidance during the summer. If they seek private tutoring, the cost is often very high. In addition, considering the time sequence of the competitions, the preparation for the AMC is often regarded as a preparatory stage and foundation for participating in subsequent competitions such as UKMT-SMC. The competition questions of AMC cover four major modules: algebra, geometry, number theory, combinatorics, and probability. They can be used to evaluate AI's problem analysis ability (RQ1) and can also effectively test AI's performance in teaching explanation and personalized guidance (RQ2). These features make AMC an ideal research object for testing whether AI can assist students in preparing for mathematics competitions.

ChatGPT-4o was chosen because it is OpenAI's flagship model and has shown high stability and maturity in mathematical reasoning and complex problem interpretation. ChatGPT-4o responds flexibly to different prompts (such as "beginner prompts"), which is convenient for experimental design and teaching scenario simulation; its API output is stable and traceable, supporting the reproducibility of research. ChatGPT has a broad application base in the field of education, and the research results have strong promotion value. In subsequent research, AI such as Gemini, Claude, and DeepSeek will also be used for comparative analysis.

This study was conducted using experimental methods and text analysis methods. For RQ1, 200 AMC questions were selected and input into ChatGPT-4o to evaluate its problem-solving correctness, knowledge point extraction and difficulty judgment, and compared with the expert annotation results. For RQ2 personalized problem-solving guidance, by setting different prompt conditions, such as "beginner prompts", we examined whether it can generate step-by-step explanations, concise language and problem-solving guidance suitable for beginners. For RQ3 personalized learning plan, ChatGPT-4o was asked to generate learning suggestions and practice plans to evaluate its overall teaching design capabilities. Representative AMC questions were selected for case analysis to demonstrate the typical output of ChatGPT-4o in problem analysis, problem-solving explanation and learning plan generation

## Findings & Discussion

### *ChatGPT-4o shows low accuracy when solving problems without prompts.*

Table 1 shows the accuracy and feasibility of ChatGPT-4o's answers under three prompting methods. Under the condition of "Question Only", the accuracy is 72% and the feasibility of the problem-solving strategy is 48%; after adding beginner prompts, the accuracy drops to 64%, but the feasibility of the idea slightly increases to 52%; when "Question + Full Solution" are provided at the same time, the accuracy reaches 100%, and the feasibility of the problem-solving strategy increases to 84%.

Prompting Strategy	Answer Accuracy(%)	Solution Feasibility (%)
Question Only	72	48
Question + Beginner Prompt	64	52
Question + Full Solution	100	84

Table 1: Performance of ChatGPT-4o under Different Prompting Strategies

The "feasibility of ideas" here refers to whether the solution given by AI conforms to the logical method commonly used by mathematics competition teachers or experienced students, focusing on the practicality of its problem-solving strategy under time-limited competitions. As shown in Figure 1, although ChatGPT-4o obtained the correct answer, it conducted multiple rounds of enumerative calculations, which is usually not recommended for students to use this problem-solving strategy in time-limited mathematics competitions.

#### Step 2: Think about possible sets of prime numbers

We need the sum of 5 different prime numbers.

Starting from the smallest primes:  
2, 3, 5, 7, 11, 13, 17, ...

Let's try summing the 5 smallest primes first:  
 $2 + 3 + 5 + 7 + 11 = 28$

Now check whether 28 is a prime:  
No, because  $28 = 2 \times 14$  ❌

Let's try replacing 11 with the next smallest prime, such as 13:

$2 + 3 + 5 + 7 + 13 = 30$  ❌

$2 + 3 + 5 + 7 + 17 = 34$  ❌

$2 + 3 + 5 + 7 + 19 = 36$  ❌

$2 + 3 + 5 + 7 + 23 = 40$  ❌

$2 + 3 + 5 + 7 + 29 = 46$  ❌

All of these are not prime numbers.

Figure 1. ChatGPT-4o's solution steps for 2024 AMC10A Problem 3

The data shows that the feasibility of ChatGPT-4o's ideas under the conditions of "providing only questions" and "questions + beginners' tips" is only 48% and 52%, which is far lower than the level of answering required to win awards in mathematics competitions (generally at least 16-18 questions out of 25 questions are answered correctly).

The study found that the feasibility of ChatGPT-4o's problem-solving strategy was greatly improved only when it provided "questions + answers", indicating that when there are reference solutions, it can better generate steps that conform to the logic of the competition. However, in the last few difficult questions of AMC, ChatGPT4o still found it difficult to give efficient and feasible problem-solving ideas, which highlights the limitations of generative AI on highly complex problems.

The reason why "question + answer" is included in the study as a prompt method is that most mathematics competitions (including AMC10) provide official answers to past exam questions. However, these answers usually only give the core ideas or key steps, which are often difficult for beginners to understand and require further explanations from teachers or experienced people. Therefore, this study pays special attention to whether AI can help beginners truly master the problem-solving ideas through clearer step-by-step analysis and concept sorting when there are already standard solutions, rather than just presenting the answers.

### *ChatGPT-4o lacks the effectiveness to solve certain types of problems*

When only the questions are provided, it can be found that many of the answers given by ChatGPT-4o to number theory, algebra, statistics and probability problems are acceptable, but its ability to solve geometry problems is limited.

#### **Problem**

One side of an equilateral triangle of height 24 lies on line  $\ell$ . A circle of radius 12 is tangent to line  $\ell$  and is externally tangent to the triangle. The area of the region exterior to the triangle and the circle and bounded by the triangle, the circle, and line  $\ell$  can be written as  $a\sqrt{b} - c\pi$ , where  $a$ ,  $b$ , and  $c$  are positive integers and  $b$  is not divisible by the square of any prime. What is  $a + b + c$ ?

- (A) 72 (B) 73 (C) 74 (D) 75 (E) 76

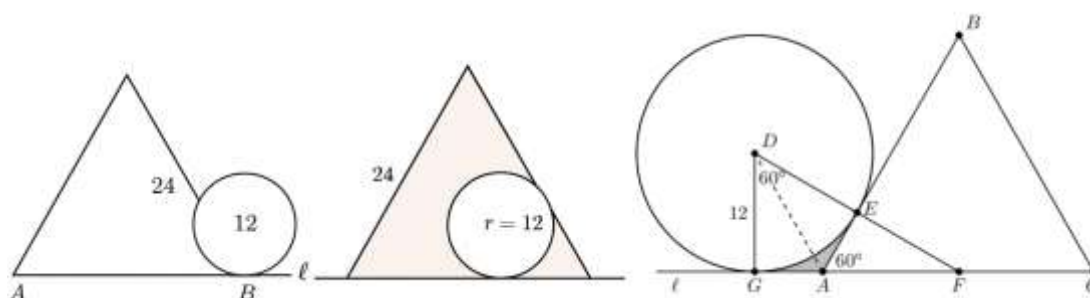


Figure 2. ChatGPT-4o's classic misunderstanding of a geometry competition question


As shown in Figure 2, this is ChatGPT-4o's understanding of the concept of "externally tangent" in a geometry competition question without prompt words. The researchers tried many times and found that ChatGPT-4o could not give the correct geometric relationship based on the description of the problem.

Comparing the figures provided in the standard solution on the right side of Figure 2, it is obvious that ChatGPT-4o cannot determine where the "area enclosed by triangles and circles" required by the question is, which is a key step in the subsequent calculation.

### *ChatGPT-4o demonstrates the ability to provide effective personalized guidance*

The study also found that ChatGPT-4o showed strong potential in personalized teaching support. Using it to classify topic questions for AMC problems, ChatGPT-4o's accuracy rate exceeded 95%, showing its practical value in teacher preparation and competition problem sorting. As shown in Figure 3, ChatGPT-4o can provide instant feedback based on specific competition problems, accurately identify the core knowledge points involved in the questions and form a relatively complete summary,

helping students build a clear knowledge system and understand the mathematical knowledge that must be learned in order to correctly solve certain problems.

 **Key Knowledge Points Summary:**

Concept	Explanation
Definition of a Prime Number	A natural number greater than 1 that has no other divisors except 1 and itself.
Odd + Odd = Even	If the sum includes the even prime 2, the result is even; without 2, the sum of five odd primes is still odd, which may be a prime.
Smallest Construction Method	Start testing combinations from the smallest possible values and replace step-by-step until the condition is met.
Digit Sum	Tests whether students understand the concept of "digit sum," i.e., summing the individual digits of a number.

Figure 3. ChatGPT-4o provides a summary of knowledge points for a competition question

AI's multilingual output capabilities, ability to provide real-time feedback at any time, and adaptive output mechanism can help address the geographical, time and resource limitations of traditional mathematics competition teaching, and provide students with more immediate and efficient mathematics competition learning support, which is also the direction of subsequent research and exploration.

## References

- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100002. <https://doi.org/10.1016/j.caeai.2020.100002>
- de Losada, M. F., & Taylor, P. J. (2022). Perspectives on mathematics competitions and their relationship with mathematics education. *ZDM—Mathematics Education*, 54(5), 941–959. <https://doi.org/10.1007/s11858-022-01404-z>
- 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>
- Mishra, P. (2019). Considering contextual knowledge: The TPACK diagram gets an upgrade. *Journal of Digital Learning in Teacher Education*, 35(2), 76–78. <https://doi.org/10.1080/21532974.2019.1588611>
- Mavrikis, M., & Margeti, M. (2024). Review of mathematics education in the age of artificial intelligence. *Research in Mathematics Education*. <https://doi.org/10.1080/14794802.2024.2389418>
- Radović, S., Marić, M., & Passey, D. (2019). Technology enhancing mathematics learning behaviours: Shifting learning goals from “producing the right answer” to “understanding how to address current and future mathematical challenges.” *Education and Information Technologies*, 24(1), 103–126. <https://doi.org/10.1007/s10639-018-9763-x>
- Voskoglou, M. G., & Salem, A. B. M. (2020). Benefits and limitations of the artificial with respect to the traditional learning of mathematics. *Mathematics*, 8(4), 611. <https://doi.org/10.3390/math8040611>
- 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>