



ChatGPT-Based Simulation Helps to Develop the Pre-Service Mathematics Teachers' Critical Thinking

Marina Drushlyak

Prof. Dr., corresponding author, Department of Mathematics, Sumy State Pedagogical University named after A. S. Makarenko, Ukraine, marydru@fizmatsspu.sumy.ua

Tetiana Lukashova

Prof. Dr., Department of Mathematics, Sumy State Pedagogical University named after A. S. Makarenko, Ukraine, tanya.lukashova2015@gmail.com

Volodymyr Shamonina

Assoc. Prof. Dr., Department of Computer Science, Sumy State Pedagogical University named after A. S. Makarenko, Ukraine, shamonawg@gmail.com

Olena Semenikhina

Prof. Dr., Department of Computer Science, Sumy State Pedagogical University named after A. S. Makarenko, Ukraine, e.semenikhina@fizmatsspu.sumy.ua

The development of IT affects society in general and the education sector in particular. The emergence of AI (in particular ChatGPT), which is constantly being improved, already causes specific problems in the organization and administration of the educational process. Still, the peculiarities of its functioning make it possible to perceive ChatGPT as learning means in the pre-service mathematics teachers' training. The article presents the results of a study that answers the following questions: (1) "Can ChatGPT be a digital platform that generates pupils' answers (secondary school)? - Yes"; (2) "Are students (pre-service mathematics teachers) prepared to check answers that ChatGPT generates? - Yes"; (3) "Does checking answers that ChatGPT generates affect the development of students' critical thinking? - Yes". The study involved a pedagogical experiment in which authors described how ChatGPT generates solutions to math problems and what mistakes it makes. We ensured that students sometimes find mathematical mistakes in the answers generated by ChatGPT, but repeated use of ChatGPT shows its positive impact on the development of pre-service mathematics teachers' critical thinking.

Keywords: ChatGPT, pre-service mathematics teachers, critical thinking, math mistakes, ChatGPT-based simulation, artificial intelligence, education

INTRODUCTION

Today, the education sector is actively use information technologies, the implementation of which is discussed on various digital platforms and analyzed in more

Citation: Drushlyak, M., Lukashova, T., Shamonina, V., & Semenikhina, O. (2025). ChatGPT-based simulation helps to develop the pre-service mathematics teachers' critical thinking. *International Journal of Instruction*, 18(1), 153-172.

depth by the academic community. Recently, researchers have noted the popularity of artificial intelligence (AI) in education, which is developing rapidly but has yet to be sufficiently studied at the level of educational practices. It has already led to some challenges (Stoker-Walker, 2022; Zhai, 2023) and questions, such as how to distinguish reliable information from deception and misinformation (Rusandi et al., 2023); how to ethically evaluate work done by AI-generated rather than by students (Mhlanda, 2023); what potential implications the evaluation of AI-generated work may have (Rudolph, 2023); how AI will affect scientific integrity (Cotton, 2023; Shiri, 2023).

Scientists also touch on the problem of developing students' thinking skills (Susjak, 2022). They are one of the most frequently mentioned competencies considered essential for academic and career success, and they play a central role in logical thinking, decision-making, and problem-solving (Altun & Yildirim, 2023). Critical thinking is essential for all professionals, without exception. Its presence is significant in the work of teachers, as every day, teachers have to check pupils' works, evaluate answers, point out mistakes, and correct them (with reasoning) (Palavan, 2020). As scientists continue to record the widespread and active use of IT in education, the issue of finding positive practices in the use of AI and the development of pre-service teachers' critical thinking are issues that require comprehensive research. An analysis of existing practices shows, in particular, that attention is paid to developing critical thinking skills among students for the effective use of AI (Rusandi et al., 2023). At the same time, we did not find any scientific results that would describe the use of ChatGPT as means for organizing quasi-professional practice (as simulation-based learning (Ören et al., 2017) for pre-service mathematics teachers.

Research Goal and Questions

Given the limited number of studies that have examined the effects of using the ChatGPT to prepare mathematics teachers, this study will examine this avenue. To verify this, we try to answer the following research questions.

Research question 1 – Can ChatGPT be a digital platform that generates pupils' answers (secondary school)?

Research question 2 – Are students (pre-service mathematics teachers) prepared to check answers that ChatGPT generates?

Research question 3 – Does checking answers that ChatGPT generates affect the development of students' critical thinking?

Literature Review

Possibilities of using ChatGPT as learning means

At the beginning of 2023, GPT (Generative Pre-training Transformer) gained popularity. It is a type of Artificial Intelligence (AI) that utilizes machine learning algorithms to generate text in natural language. Starting from February 18, 2023, this service became available in Ukraine, and it began to be used in the educational process not only worldwide (Adipat, 2023; AlKanaan, 2022; Elgohary & Al-Dossary, 2023; Nurhasan et al., 2022) but also in Ukraine (Baibakova & Hasko, 2023; Golub et al.,

2023; Melnyk, 2023; Podlasov & Matviichuk, 2023; Shulzhenko et al., 2023; Shyshkina & Nosenko, 2023; Sisilitsyn & Osadchyi, 2023; Yandola, 2023).

Developers of ChatGPT envisioned various capabilities, including text completion, essay, and story, social media post writing, summarization, classification, paraphrasing, translation, question answering (providing medical advice, expressing opinions on subjective topics), composing emails, generating code in different programming languages, and analyzing communication tone (Nanovska, 2023). Ideas for using ChatGPT as a language learning tool (Kohnke et al., 2023; Nugroho et al., 2023), programming (Hartley et al., 2024), flipped learning means (Li, 2023) are proposed.

We review several studies on using ChatGPT as math learning means. O. Taani and S. Alabidi (2024) see ChatGPT-using in generating examples, assessing difficulty, providing explanations, supporting problem-solving, and test preparation. Other researchers argue that “the use of ChatGPT as a teaching tool in mathematics instruction has the potential to enhance student performance, but the role of the instructor remains crucial in delivering in-depth instruction. While ChatGPT can provide information, explanations, and support, the combination of human expertise and ChatGPT’s capabilities holds greater potential for improving students’ understanding of mathematics” (Dadan et al., 2024).

However, as highlighted in the findings (Gao et al., 2023), ChatGPT needs an understanding of the meanings of the words it processes. While AI recognizes patterns and generates plausible responses, it does not comprehend the concepts it operates with (Bogost, 2022). ChatGPT provides responses likely based on learned patterns, which could be a source of potential mistakes or inaccuracies. It is important to note that ChatGPT is "dependent" on the correctness of the prompt formulation, available hints, or clarifications. The mentioned aspects and the authors' experience in "communicating" with ChatGPT have inspired using ChatGPT as a digital tool in mathematics teachers' preparation.

ChatGPT for the development of mathematics teachers’ critical thinking

Critical thinking, as a specific type of thinking, is characterized by the ability to identify flaws in reasoning or conclusions. An analysis of definitions of critical thinking indicates that it is an active, thoughtful, and purposeful process that occurs during the interpretation and evaluation of information and experience to clarify or enhance understanding of something (Pnevmatikos et al., 2023; Altun & Yildirim, 2023). Such a process is only possible with the skills to analyze facts, evaluate arguments, justify positions, make comparisons, and predict consequences (Arisoy & Aybek, 2021) and the ability to interpret arguments and draw conclusions to support one's judgments (Facione, 2015).

As (Kozachenko, 2017) points out, critical thinking is only possible with an intellectual foundation, which includes a sufficient level of general and specific knowledge, mastery of basic methods of cognitive activity and logical argumentation, and proficiency in effective problem-solving techniques. Applying these principles to the professional training of a mathematics teacher, it becomes clear that the development of

critical thinking in mathematics learning typically occurs when solving mathematical (computational, logical, graphical, etc.) problems or through verification of solutions from others. AI can generate solutions that can be used as school student-answer simulations, and, therefore, it can be a digital tool in professional development.

ChatGPT-based simulation in pre-service teachers' preparation

Simulation-based learning replicates certain aspects of reality to shape and develop specific professional skills (Cruz & Patterson, 2005), for example, in solving typical tasks in a particular field. Simulation in the educational process is considered a teaching method that facilitates the acquisition of target complex skills (Cook, 2014). The effectiveness of simulation is associated with the ability to create diverse situations that allow the acquisition of quasi-professional experience (Theelen et al., 2019). Compared to traditional teaching, simulation has more potential for developing practical skills, generating ideas, and demonstrating abilities (Theelen et al., 2019; Chernikova et al., 2020). Using simulations improves the understanding of acquired knowledge and skills, identifies gaps in the system of professional skills, and allows addressing them during the professional training stage (Levin & Flavian, 2020). In particular, the research conducted by B. Banić et al. (2023) has demonstrated that ChatGPT can serve as a simulation tool in training future programmers. The generalization of these results suggests the possibility and potential usefulness of using AI in math teacher training. ChatGPT can be a virtual environment that proposes the simulation of problem-solving and further solution analysis, forming the basis for developing students' critical thinking.

METHOD

Research Methodology

Participants

The study was conducted at Sumy State Pedagogical University in 2023. The experiment involved students majoring in secondary education (mathematics). The total number of participants was 32, including 14 3rd year math students of the Bachelor's degree and 18 1st year math students of the Master's degree. The students' ages ranged from 22 to 28, with an average age of 22.9. Most of the participants in the experiment were girls (26 people, 81%). Participation in the experiment was voluntary. All participants agreed to participate in the pedagogical experiment. The experiment took place as part of the study of the educational components "Teaching Mathematics with Computer Support" (Bachelor) and "School Mathematics Course with Computer Support" (Master).

Research design

A mathematics teacher's preparation involves creating conditions close to the real ones. ChatGPT can generate answers to any questions, so it can be considered an environment that acts as a pupil answering questions. In other words, we perceive ChatGPT as a simulator of answers given by pupils studying at school. The idea of generating answers (correct and not entirely accurate, although plausible) became the basis for creating the experiment.

Research Question 1

We planned to find the answer to the first question by working directly with ChatGPT in the "prompt-response" format. ChatGPT generates statistically most likely answers (that features of the algorithms (Dadan et al., 2024). The answers are inaccurate sometimes. Therefore, these answers can be perceived as the pupils' answers (secondary school).

Research Question 2

The answer to the second question was based on the following considerations. The job of a mathematics teacher involves checking test papers. Pre-service teachers need to learn to "see" pupils' mistakes, follow pupils' reasoning, and notice inconsistencies in their thinking. We used the ChatGPT answers to check whether pre-service mathematics teachers could find mathematical mistakes and false statements in ChatGPT answers.

During one of the classes with pre-service mathematics teachers (the topic "Methods of teaching the method of mathematical induction"), we explained in detail how to use ChatGPT to solve problems and analyzed the generated answers in detail (more on this in section 3.1). After the class, we asked pre-service mathematics teachers to complete an individual work "Using the Method of Mathematical Induction" (Table. 1). This was to demonstrate their ability to check "pupils' work" (the ChatGPT played the role of a secondary school pupil and generated solutions).

Table 1

Tasks of individual work (Source: made by authors)

Task 1.	Propose to ChatGPT to prove the inequality. What proof method does ChatGPT prioritize? What methods does it suggest for proving? How many of them? Record the results for the report.
Task 2	Identify any mistakes made by ChatGPT (if any). Formulate them for the report.
Task 3	Point out ChatGPT's mistakes. Analyze whether ChatGPT corrected the solution. Did it take into account your previous comments? Record the results for the report.
Task 4	Propose to ChatGPT to prove the given inequality using different methods (at least two other methods). Are there any mistakes in these methods? Record the results for the report

We subjected the results to a qualitative analysis: did the pre-service mathematics teachers find any incorrect answers generated by ChatGPT? Were the pre-service mathematics teachers able to pinpoint the mistake correctly and eventually get the correct answer? If pre-service mathematics teachers found at least one inaccuracy in the generated answer and, after asking the ChatGPT, came up with the proper result, we considered them able to check pupils' answers to math tasks. It took up to 2 hours to complete the tasks

Research Question 3

Learning activities that involve analyzing answers are always associated with critical thinking. Therefore, we assumed that working with the ChatGPT to solve mathematical tasks can influence pre-service mathematics teachers' critical thinking. Thus, the

ChatGPT can become a digital tool for developing critical thinking. To test this assumption, we had to decide on indicators: what can we use to test the development of critical thinking? After analyzing the scientific literature, we concluded that the indicators could be (1) Ability to analyze information critically, (2) Ability to identify logical inconsistencies in statements, (3) Ability to rectify logical inconsistencies, (4) Tendency to seek the most rational approach to problem-solving.

We also realized that pre-service mathematics teachers' critical thinking development occurs over time, so we needed to measure each student's critical thinking level twice. Therefore, we organized a pedagogical experiment in two stages: the first measurement of the pre-service mathematics teachers' critical thinking level was based on the results of their work at once, and the second measurement was based on the execution results of the same individual work but a month later.

We linked the indicators of critical thinking development to the tasks of individual work.

- Indicator 1: The ability to analyze information critically. When a pre-service mathematics teacher successfully traces a mathematical induction algorithm step by step, it's a testament to his ability to analyze information critically. This indicator corresponds to Task 1.
- Indicator 2: The ability to identify logical inconsistencies in statements. If pre-service mathematics teachers can find at least one mistake or logical inconsistency in the proposed solution, they have the ability to identify logical inconsistencies in statements. This indicator corresponded to Task 2.
- Indicator 3: The ability to rectify logical inconsistencies. If pre-service mathematics teachers can correct logical inconsistencies by providing arguments and pointing out mistakes, they have the ability to rectify logical inconsistencies. This indicator corresponds to Task 3.
- Indicator 4: Tendency to seek the most rational approach to problem-solving. That is a crucial aspect of critical thinking. When a pre-service mathematics teacher can create a request to find alternative ways to solve a problem and can evaluate them critically, it's a strong indication of their ability to seek the most rational approach to problem-solving. This indicator corresponded to Task 4.

The peculiarities of assessing individual work tasks are shown in Table 2

Table 2
Features of assessment (Source: made by authors)

Score	Indicator 1	Indicator 2	Indicator 3	Indicator 4
0	The student does not differentiate the steps of the induction method	The student does not identify mistakes	The student does not formulate arguments regarding mistake rectification	The search was conducted without analyzing the provided results
1	-	The student identifies only one mistake	The student incorrectly formulates arguments for mistakes	The search for the most rational approach was conducted but not supported by a reasoned position
2	The student distinguishes the steps of the induction method	The student identifies more than one mistake	The student formulates arguments for mistake rectification and achieves a correct result	The search for the most rational approach was conducted, and the own position was argued

Visualization of the experiment is shown in Fig. 1.

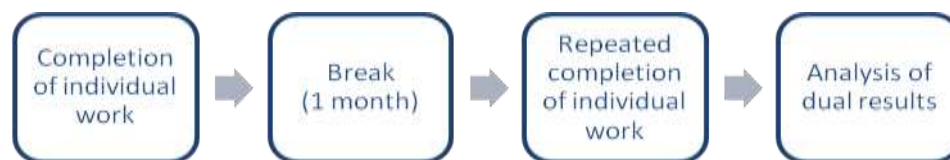


Figure 1
Experiment design (source: made by authors)

The task's results were evaluated using 0 to 2 points each. The maximum score for a paper was 8 points. We analyzed the experiment's results in pairs (each pre-service mathematics teacher submitted two papers, with a month between them). The results were independent between pairs but dependent within pairs. It was the interval measurement scale. This experiment design allowed us to use the non-parametric Wilcoxon test to determine the significance of differences in pre-service mathematics teachers' critical thinking development.

We formulate two hypotheses: " H_0 : median $D_i \leq 0$ (solving tasks in the ChatGPT environment doesn't improve pre-service mathematics teachers' critical thinking)" and

" H_a : median $D_i > 0$ (solving tasks in the ChatGPT environment improves pre-service mathematics teachers' critical thinking)".

FINDINGS

ChatGPT as a digital platform for generating pupils' answers (research question 1)

We asked ChatGPT several math questions and saw answers that were not always correct. The reasoning presented was only sometimes correct. ChatGPT generated different answers on different computers. That affirmed that the answers from ChatGPT could be pupils' answers. We describe our experience below in more detail.

To solve Task 1, we formulate the prompt "Prove the inequality $2^n - n^3 > 23$ for all natural $n \geq 10$ " in the ChatGPT environment. ChatGPT generates a response (Fig. 2).

You
Довести нерівність $2^n - n^3 > 23$ для довільного $n \geq 10$.

ChatGPT
Ну давай подивимося. Спробуємо довести цю нерівність методом математичної індукції. Спочатку з базового випадку, де $n = 10$.

1. Перевіримо базовий випадок:
Для $n = 10$: $2^{10} - 10^3 = 1024 - 1000 = 24 > 23$.

Отже, базовий випадок виконується.

2. Припустимо, що для деякого k нерівність виконується, тобто:
 $2^k - k^3 > 23$.

3. Доведемо, що і для $k + 1$ нерівність також виконується:
 $2^{k+1} - (k+1)^3 = 2 \cdot 2^k - (k^3 + 3k^2 + 3k + 1)$
Ми вже знаємо, що $2^k - k^3 > 23$, тому покладемо на решту виразу:
 $-(k+1)^3$
 $(k+1)^3 = k^3 + 3k^2 + 3k + 1$ ↓

Підставимо це у вираз:
 $2^{k+1} - (k+1)^3 = 2 \cdot 2^k - (k^3 + 3k^2 + 3k + 1)$
 $2^{k+1} - (k+1)^3 = 2^k - k^3 + 2^k - 3k^2 - 3k - 1$
Оскільки ми вже припустили, що $2^k - k^3 > 23$, можна записати це так:
 $2^{k+1} - (k+1)^3 > 2^k - k^3 + 2^k - 3k^2 - 3k - 1 > 23 - 3k^2 - 3k - 1$.

Нині залишається показати, що $23 - 3k^2 - 3k - 1 > 23$.

Після спрощення отримуємо:
 $-3k^2 - 3k - 1 > 0$.

Ця завжди виконується для будь-якого $k \geq 10$, оскільки коефіцієнти перед k^2 та k негативні, а константа додатна.

Отже, за допомогою методу математичної індукції ми показали, що нерівність $2^n - n^3 > 23$ виконується для будь-якого $n \geq 10$.

Figure 2

The result of completing the Task 1 (source: made by authors)

The ChatGPT chooses the method of mathematical induction, presumably due to the requirement in the problem statement "to prove for all natural numbers n ". This method is one of the most common ways to prove such inequalities, so the chatbot selected a standard reasoning approach. It is worth noting that ChatGPT outlines all the steps of the mathematical induction method and provides what appears to be a correct proof, but

in the highlighted line after using the induction assumption, it omits the term 2^k , which plays a crucial role in further evaluations of the expression.

Furthermore, in the provided proof, it is claimed that $-3k^2 - 3k - 1 > 0$ for all $k \geq 10$, which is not true. Therefore, analyzing the generated proof is crucial for identifying mistakes and exploring ways to rectify them.

We formulate the following prompt to specify the mistake: "The mistake occurred during the induction step." ChatGPT attempts to respond to the specified mistake with a new prompt and correct it, but it fails to do so. It essentially proposes the same proof (Figure 3).

We formulate the prompt, indicating the mistake more specifically "There is still the

mistake when using the inductive assumption $2^k - k^3 > 23$ wasted 2^k in transformations and an unreasonable conclusion was made». After being pointed out more specifically its mistake, ChatGPT suggests the same proof with an incorrect

subtraction of 1 for compensation, losing the term 2^k again. In other words, ChatGPT cannot correct the mistake even with external assistance; it reverts to the previous proof (Figure 4).



Figure 3
The reaction (1) of ChatGPT to pointing out its mistakes (source: made by authors)



Figure 4
The reaction (2) of ChatGPT to pointing out its mistakes (source: made by authors)

Rephrase the prompt, indicating the mistake with even more precision "This does not fix the situation, the additive is lost 2^k , which exceeds the value of the trinomial $3k^2 + 3k + 1$ at values $k \geq 10$ and allows to make a correct assessment». Specify the mistake in the prompt and get proof that is not complete but does not contain logical and factual mistakes (Figure 5). To complete the proof, it remains to show that for $k \geq 10$, $2^k - 3k^2 - 3k - 1 > 0$. Note that the result of the proof of the same inequality proposed by ChatGPT after a month also contains mistakes (Prompt «Prove the inequality $2^n - n^3 > 23$ for all-natural $n \geq 10$ », Figure 6). Despite the more detailed proof, ChatGPT makes the same mistakes in transformations.

To solve Task 4, we ask ChatGPT to prove the given inequality using different methods (Prompt "Use other methods to prove the inequality"). ChatGPT can generate other methods of proving the given inequality— "the method of using functions", "the method of using function analysis", "the differentiation method", "the method of reduced factors" (keeping the names of the methods as suggested by ChatGPT).

However, "proofs" contain either obvious mistakes in transformations or "hide" them in a correct reasoning shell. None of the proofs generated by ChatGPT were correct. For example, the "proof" (method of using functions, Figure 7) contains many mistakes, starting from the estimation at $n=10$, with an arithmetic mistake in finding 2^{10} , to the false statement that $f(n+1)-f(n)$ should be less than or equal to zero.

Another method suggested by ChatGPT (Method of using function analysis, Figure 8) also has several drawbacks – it provides general recommendations for proving inequalities using the monotonicity of functions without specifying how to implement it in this particular case.



Figure 5
The reaction (3) ChatGPT to mistake correction (source: made by authors)



Figure 6
The result of the task solution after a month (source: made by authors)



Figure 7
The method of using a function (source: made by authors)



Figure 8
Method of Using Function Analysis (source: made by authors)

When using the method of differentiation (Figure 9), ChatGPT applies the assumption of induction, incorrectly finds the derivative, and uses the differentiation of inequality, which is generally incorrect.

When using the method of shortening factors (Figure 10), an attempt is made to use factorization formulas, but this attempt is not implemented. Then, ChatGPT limits itself to general recommendations for proving the inequality by the same functional method using the derivative (mistakenly assuming that at $n=10$ there will be a minimum (actually, the minimum value of the function) without a specific implementation).

Метод диференціювання:

- Базовий випадок:** Перевіримо для $n = 10$:
 $2^{10} - 10^3 = 1024 - 1000 = 24 > 23$.
- Індукційний крок:** Припустимо, що нерівність виконується для певного k : $2^k - k^3 > 23$.
 Тепер диференціюємо обидві сторони нерівності за n :
 $2^{k+1} - 3k^2 > 0$.
 Тут ми використали припущення, що $2^k - k^3 > 23$ для обчислення похідної $2^{k+1} - (k+1)^3$.
 Довелимо, що сторона нерівності \uparrow зростає для $k + 1$.

Figure 9
Differentiation Method (source: made by authors)

Метод скорочення множників:

- Розглянемо нерівність: $2^n - n^3 > 23$.
- Розкриємо різницю квадратів: $2^n - n^3 - 23 > 0$.
 $(2^{n/2})^2 - n^3 + 23 > 0$.
- Розглянемо функцію: $f(n) = (2^{n/2})^2 - n^3 + 23$.
- Покажемо, що $f(n)$ має мінімум для $n \geq 10$.
 * Розглянемо $f'(n)$ (знайдемо точку, де $f'(n) = 0$)
 * Проведемо тестування за допомогою другої похідної для підтвердження, що це мінімум.
 * Перевіримо значення $f(n)$ в точці мінімуму.

Figure 10
Method of reduced factors (source: made by authors)

Therefore, ChatGPT makes mistakes in proving inequalities, providing an opportunity to use it for generating student-like responses. Generated solutions containing mistakes become important training tools in teacher preparation in general and for developing students' critical thinking tools in particular. It is worth noting that the response of ChatGPT depends on the formulation of the task itself, and the wording of clarifications or comments on the proposed solutions. Additionally, ChatGPT responses are influenced by the language of the prompt (specifically, the generation of solutions is more accurate in English). So, all the results were presented in the Ukrainian language, the language in which the experiment was conducted. Note, that when the prompt is formulated in English, ChatGPT provides correct proof without mistakes, although without detailed explanations in the final steps (Figure 11).

You
 Prove the inequality $2^n > n^3 > 23$ for an arbitrary natural $n \geq 10$

ChatGPT
 Certainly, let's prove the inequality $2^n - n^3 > 23$ for any natural $n \geq 10$.

Proof by Mathematical Induction:

Base Case:
 For $n = 10$, let's verify the inequality:
 $2^{10} - 10^3 = 1024 - 1000 = 24 > 23$.

The base case holds true.

Inductive Step:
 Assume that for some $k \geq 10$, the inequality $2^k - k^3 > 23$ is true.

Now, let's prove for $k + 1$:
 $2^{k+1} - (k+1)^3 = 2 \cdot 2^k - (k^3 + 3k^2 + 3k + 1)$.

Consider the expression:
 $2^k - k^3 + 2^k - 3k^2 - 3k - 1$.

Now, split it into two terms:
 $2^k - k^3 + (2^k - 3k^2 - 3k - 1)$.

By the inductive assumption, the first term is greater than 23:
 $> 23 + (2^k - 3k^2 - 3k - 1)$.

Now, we need to show that $(2^k - 3k^2 - 3k - 1) > 0$. This is true for $k \geq 10$ because the exponential term 2^k grows faster than the quadratic and linear terms.

Therefore, the inequality holds for $k + 1$.

By mathematical induction, the given inequality $2^n - n^3 > 23$ is proven for all $n \geq 10$.

Figure 11
The prompt is formulated in English (source: made by authors)

Are students (pre-service mathematics teachers) prepared to check answers that ChatGPT generates? (research question 2)?

The results of the individual work are presented in Table 3, where the first individual work results are labeled as IW-1 and the second as IW-2. We did not receive zero marks for the results of the first individual work, which proved a positive answer to the second question of our study: pre-service mathematics teachers can see mistakes, and therefore, they can be considered prepared to check "pupils' answers".

Table 3
Double assessment of student works (Source: made by authors)

No	IW-1	IW-2	No	IW-1	IW-2	No	IW-1	IW-2
1	2	2	12	6	8	23	3	4
2	6	7	13	4	3	24	4	7
3	4	5	14	3	2	25	2	2
4	4	5	15	2	4	26	6	7
5	5	5	16	2	4	27	4	5
6	3	3	17	2	2	28	4	5
7	3	4	18	4	4	29	5	5
8	2	4	19	5	7	30	3	3
9	2	2	20	6	8	31	3	4
10	4	4	21	4	3	32	2	4
11	5	7	22	3	2			

There were no problems with analyzing the results of the mathematical induction algorithm (Task 1 in the individual work). It is worth noting that ChatGPT made a mistake when checking the induction base in one case. Most pre-service mathematics teachers easily coped with detecting mistakes in the transformations (Task 2). Several proposed proofs did not use the induction assumption, which led to an incorrect conclusion. Difficulties were observed in Task 3 when it was necessary to point out the mistake. If the prompt was formulated briefly, without detailed argumentation, ChatGPT did not "take it into account" and continued to offer proofs with mistakes. The most significant difficulties arose in Task 4 solving, as other proof methods offered by ChatGPT were based on knowledge from different mathematics areas, which required students to have a reasonably thorough understanding of higher mathematics to analyze them.

Does checking answers that ChatGPT generates affect the development of students' critical thinking? (research question 3)

To answer the third question of our study, we statistically analyzed the results of two individual stages using the non-parametric Wilcoxon test (King & Eckersley, 2019). A comparison of the individual work results showed that ten results remained unchanged. Other pairs showed changes, in particular, with 18 positive changes. Ranking by the Wilcoxon test gave an empirical value of $T=243$. The critical value of the test for the significance level of 0.05 and for the case of a one-sided test gave the values $W_{\alpha}=75.99$ and $W_{1-\alpha}=177$. According to the decision rule for the one-sided Wilcoxon test, we have $T > W_{1-\alpha}$. Therefore, the null hypothesis is rejected and the alternative one (solving tasks in the ChatGPT environment improves pre-service mathematics teachers' critical thinking) is accepted.

It is worth noting that 22 pre-service mathematics teachers changed their initial score, and 18 showed an increase in the final score (9 students received +1 point in the second performance; 8 students received +2 points; 1 student improved the result by +3 points). That indicates that pre-service mathematics teachers have become more attentive to the answers; they see more mistakes and inconsistencies and can write a prompt so that ChatGPT corrects the mistakes found and comes to the correct result. When we talked

to these students, we noted their interest in working with ChatGPT and their desire to "catch" the answer generator on a mistake. Four pairs of answers showed a decrease in the result. That can be explained by pre-service mathematics teachers' inattention when doing the same work for the second time. Ten pairs of works had the same marks. When we talked to the students who did these papers, we found out that they either remembered the first performance well and reproduced it (6 students) or, when performing tasks, they did not try to do them as well as possible due to a lack of encouragement (4 students).

DISCUSSION

Research Question 1. (Can ChatGPT be a digital platform that generates pupils' answers (secondary school)? In society and within the academic community, concerns have arisen regarding the use of AI ("job killer", "influences people's thoughts through generating posts and comments for social media", "generates academic texts that are not substantiated", etc.). Due to such concerns, some schools have even prohibited the use of ChatGPT (Dibble, 2023; Lukpat, 2023). For instance, S. Marche (2022) suggests that it may take "ten years for scholarly circles to confront this new reality: two years for students to understand the technology, another three years for professors to acknowledge that students are using this technology, and then five years for university administrators to decide what to do with it if anything". W. M. Lim et al. (2023) argue that, on the one hand, AI leads to the disruption of the education system, while, on the other hand, it opens up a new era of accessible information and automation to enhance the quality of education.

M. Farrokhnia et al. (2023), based on the results of the SWOT analysis, identified the strengths of ChatGPT (creating plausible responses, self-improvement or self-learning ability, providing personalized responses, real-time response capabilities, especially for education, increasing information accessibility, facilitating individual learning, supporting comprehensive learning, reducing teachers' workload, e.g., ChatGPT can create tests for you) and weaknesses (lack of deep understanding, difficulties in assessing the quality of responses, democratization of plagiarism in education/science, reduction of high-level cognitive skills such as creativity, critical thinking, reasoning, and problem-solving). Our research demonstrated the positive aspects of using AI in the educational process. ChatGPT has become a digital tool for developing critical thinking in youth by generating plausible but, from a mathematical perspective, incorrect statements, formulations, and conclusions.

Today, there is also a discussion about AI's ability to assess data accurately: ChatGPT lacks the human ability to evaluate the reliability of the data it has been trained for (Lecler et al., 2023). ChatGPT does not have access to the internet and currently has limited knowledge of events in the world after 2021 (Stokel-Walker & van Noorden, 2023), leading to the possibility of providing outdated and inaccurate answers. ChatGPT is not afraid to be "wrong"; it is afraid to be "uninformed". If it doesn't know the answer, it will "make it up" and still respond. That can be used as material for critical analysis and the development of student's critical thinking, as confirmed by our research.

Research Question 2 (Are students (pre-service mathematics teachers) prepared to check answers that ChatGPT generates)? Pre-service mathematics teachers should develop the "teach to prove mathematical statements" skills, which include not only the ability to prove theorems/equations/inequalities from the school math course but also the ability to anticipate common mistakes made by pupils, justify mistakes in pupils' work, and correctly and logically correct the mistakes made by pupils. Such skills, on the one hand, are professional for a teacher. Still, on the other hand, they characterize the level of development of their critical thinking, which needs to be enhanced, especially in situations close to real ones. Similar conditions, as our research has shown, can be generated using AI, particularly ChatGPT.

The use of ChatGPT in education calls for reconsidering assessment methods (Dwivedi, 2023). For instance, we should replace descriptive tasks (like essays) with more creative ones that require students' critical thinking and focused creative efforts. In our research, we modified typical tasks, shifting from "prove the inequality" and "identify possible mistakes in students' work" to "analyze the given response" and "detect/justify/correct mistakes." These tasks are not algorithmic and carry significant heuristic value. Our assignments generally involve abstract thinking and foster the development of skills in analysis, logical reasoning, and concluding. They also provide an opportunity to comprehensively reinforce a wide range of theoretical concepts studied in the school mathematics curriculum, including the basics of inequality theory, the equivalence of transformations, properties of functions, applications of derivatives and integrals, geometric reasoning, estimation of areas and volumes, etc.

Research Question 3 (Does checking answers that ChatGPT generates affect the development of students' critical thinking?) Modern education is concerned not only with acquiring knowledge and skills in a specific field of knowledge or future profession (hard skills). With the development of information technologies, qualities that ensure an individual's competitiveness in the modern job market have become essential. These qualities are referred to as soft skills and include analytical thinking and innovation, active learning and learning strategies, complex problem solving, critical thinking and analysis, creativity, originality and initiative, leadership and social influence, resilience, stress tolerance and flexibility, reasoning, problem-solving, and ideation (World Economic Forum, 2020), collectively known as the "Four Cs" - Collaboration, Communication, Creativity, and Critical thinking. Researchers have recently emphasized the importance of 21st-century skills (Thornhill-Miller et al., 2023). These skills are crucial for teachers, who must be creative in their professional activities and capable of facilitating communication and promoting student teamwork. Given the nature of their profession (teaching mathematics), possessing and developing critical thinking is particularly essential for mathematics teachers. Our research indicates the potential for developing math students' critical thinking as part of their professional training, contributing to cultivating the in-demand skills of the 21st century.

Our study aligns with the key ideas of the Concept for the Development of Artificial Intelligence in Ukraine (2020). Specifically, for the higher education sector, the Concept's authors emphasize the importance of incorporating artificial intelligence

topics into educational programs across various specialties. We see the implementation of this idea in AI integration as a teaching tool in the professional training of teachers – creating simulation cases with answer generation not only introduces the possibilities of using artificial intelligence and the peculiarities of answer generation but also facilitates additional quasi-professional practices.

LIMITATION

We should note the limitations of our study.

1. The experimental study was conducted on 32 pre-service mathematics teachers. We monitored critical thinking (as a personal ability). The experiment's design involved tracking this quality for each student. If the number of participants in the experiment is increased, we can obtain more reliable results.
2. In 2023, we conducted a pedagogical experiment using ChatGPT. It's important to note that AI models have significantly evolved since then, potentially impacting the practical significance of our study's findings in the present context.
3. During the pedagogical experiment, we focused on generating prompts and answers related to one of the topics of elementary mathematics, "Method of mathematical induction", which also limits the practical significance of the findings.
4. On the one hand, ChatGPT gives the university teacher a tool for developing students' critical thinking, but on the other hand, the university teacher gives the student a tool for solving educational tasks. Therefore, using ChatGP in the classes will require teachers to reconsider the traditional approach to grading student work. Accordingly, the use of ChatGPT should be conscious and balanced.
5. The experiment involving communication with ChatGPT was conducted in Ukrainian. It's worth noting that the language of the prompt influences ChatGPT responses, which means that the results could vary if the experiment were conducted in other languages.

These limitations do not affect the general conclusions of our study on the feasibility of using ChatGPT to prepare pre-service mathematics teachers.

CONCLUSIONS

The new educational trend involves implementing artificial intelligence in the educational process. ChatGPT can be considered a means of simulating interactions with pupils in pre-service mathematics teachers' preparation. Our study shows that the ChatGPT environment can be a digital platform that generates students.

Math students should not only know how to prove the theorems of the school mathematics course but also be able to predict typical mistakes that students may make in their proofs. Our study showed that math students can check answers that ChatGPT generates.

A high level of critical thinking is necessary for working with artificial intelligence. Proving mathematical statements is relevant material for developing critical thinking.

The study results confirmed the effectiveness of using ChatGPT in developing the critical thinking of math students.

So, artificial intelligence can be considered a tool for pre-service math teachers preparing. However, integrating ChatGPT in education presents both opportunities and challenges for teachers. By understanding ChatGPT's capabilities and limitations, math students can use it more effectively in future professional activities.

REFERENCES

- Adipat, S. (2023). An artificial intelligence-enhanced phenomenon-based learning approach for interdisciplinary understanding and speaking skills. *International Journal of Instruction*, 16(3), 531-550. <https://doi.org/10.29333/iji.2023.16329a>.
- AlKanaan, H. M. N. (2022). Awareness regarding the implication of artificial intelligence in science education among pre-service science teachers. *International Journal of Instruction*, 15(3), 895- 912. <https://doi.org/10.29333/iji.2022.15348a>.
- Altun, E. & Yildirim N. (2023). What does critical thinking mean? Examination of pre-service teachers' cognitive structures and definitions for critical thinking. *Thinking Skills and Creativity*, 49, 101367. <https://doi.org/10.1016/j.tsc.2023.101367>.
- Arisoy, B. & Aybek, B. (2021). The Effects of Subject-Based Critical Thinking Education in Mathematics on Students' Critical Thinking Skills and Virtues. *Eurasian Journal of Educational Research*, 92, 99-120. <https://doi.org/10.14689/ejer.2021.92.6>.
- Baibakova, I. M. & Hasko, O. L. (2023). Pros and cons of ChatGTP language model in educational process. *Innovative Pedagogy*, 64 (1), 104-107. <https://doi.org/10.32782/2663-6085/2023/64.1.20>.
- Banić, B., Konecki, M. & Konecki, M. (2023). Pair Programming Education Aided by ChatGPT. In MIPRO 2023: 46th ICT and Electronics Convention, Opatija, Croatia, 981-085.
- Bogost, I. (2022). ChatGPT is dumber than you think. *The Atlantic*. Retrieved Jan 03, 2024 from <https://www.theatlantic.com/technology/archive/2022/12/chatgpt-openai-artificial-intelligence-writing-ethics/672386/>.
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, 90(4), 499-541. <https://doi.org/10.3102/0034654320933544>.
- Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education*, 48(8), 750-760. <https://doi.org/10.1111/medu.12473>.
- Cotton, D. R., Cotton, P. A. & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12. <https://doi.org/10.1080/14703297.2023.2190148>.

- Cruz, B. C. & Patterson, J. (2005). Cross-Cultural Simulations in Teacher Education: Developing Empathy and Understanding. *Multicultural Perspectives*, 7(2), 40-47. https://doi.org/10.1207/s15327892mcp0702_7.
- Dadan D., Hendriyanto A., Sahara S., Suryadi D., Muhaimin Lukman H., Chao T. & Fitriana L. (2024). ChatGPT in a didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. *Frontiers in Education*, 8. <https://doi.org/10.3389/educ.2023.1295413>.
- Dibble, M. (2023). Schools ban ChatGPT amid fears of artificial intelligence-assisted cheating. *VOA News*. Retrieved 25 February, 2023 from <https://www.voanews.com/a/schools-banchatgpt-amid-fears-of-artificial-intelligence-assisted-cheating/6949800.html>.
- Dwivedi, Y. K. (2023). 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, 1-63. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>.
- Elgohary, H. K. A., & Al-Dossary, H. K. (2023). The effectiveness of an educational environment based on artificial intelligence techniques using virtual classrooms on training development. *International Journal of Instruction*, 15(4), 1133-1150. <https://doi.org/10.29333/iji.2022.15460a>.
- Facione, P.A. (2015). Critical thinking: What it is and why it counts. *Insight Assess*, 1, 1-23.
- Farrokhnia, M., Banihashem, S. K., Noroozi, O. & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*. <https://doi.org/10.1080/14703297.2023.2195846>.
- Gao, J., Zhao, H., Yu, C. & Xu, R. (2023). Exploring the feasibility of ChatGPT for event extraction. *arXiv*. <https://doi.org/10.48550/arXiv.2303.03836>.
- Golub, T. P., Zhygzyhtova, L. M., Kovalenko, O. O. & Nazarenko, O.I. (2023). ChatGPT and its influence on modern higher education. *Innovative Pedagogy*, 58(1), 201-204. <https://doi.org/10.32782/2663-6085/2023/58.1.40>.
- Hartley, K., Hayak, M. & Ko, U.H. (2024). Artificial Intelligence Supporting Independent Student Learning: An Evaluative Case Study of ChatGPT and Learning to Code. *Educ. Sci.*, 14, 120. <https://doi.org/10.3390/educsci14020120>.
- Kohnke, L., Moorhouse, B. L. & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>.
- Kozachenko, N. P. (2017). Critical thinking: marginal approaches and optimal ways. *Actual problems of spirituality*, 18, 165-178.
- Lecler, A., Duron L. & Soyer, P. (2023). Revolutionizing radiology with GPT-based models: Current applications, future possibilities, and limitations of ChatGPT.

Diagnostic and Interventional Imaging, 104, 6, 269-274.
<https://doi.org/10.1016/j.diii.2023.02.003>.

Levin, O. & Flavian, H. (2020). Simulation-based learning in the context of peer learning from the perspective of pre-service teachers: a case study. *European Journal of Teacher Education*. <https://doi.org/10.1080/02619768.2020.1827391>.

Li, H. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, 39(5), 40–58. <https://doi.org/10.14742/ajet.8923>.

Lim, W. M., Gunasekara, A., Pallan, J. L., Pallant, J. I. & Pechenkina E. (2023). Generative AI and the future of education: Ragnarok or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21, 100790. <https://doi.org/10.1016/j.ijme.2023.100790>.

Lukpat, A. (2023). ChatGPT banned in New York City public schools over concerns about cheating, learning development. *The Wall Street Journal*. Retrieved 27 February, 2023 from <https://www.wsj.com/articles/chatgpt-banned-in-new-york-city-public-schools-over-concerns-about-cheating-learning-development-11673024059>.

Marche, S. (2022). The college essay is dead. Nobody is prepared for how AI will transform academia. *The Atlantic*. Retrieved 25 February, 2023 from <https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371>.

Melnyk, Yu. B. (2023). Using of ChatGPT in Psychology Research and Practice. *International Journal of Science Annals*, 6, 2, 5-8. <https://doi.org/10.26697/ijsa.2023.2.0>.

Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.4354422>.

Nanovska, V. (2023). What you need to know about ChatGPT so that it becomes an effective assistant and not a trap. *Mediamaker*. <https://mediamaker.me/chat-bot-chatgpt-shho-potribno-znaty-vydavczyam-569>.

Nugroho, A., Putro, N. H. P. S. & Syamsi, K. (2023). The Potentials of ChatGPT for Language Learning: Unpacking its Benefits and Limitations. *Register Journal*, 16(2), 224–247. <https://doi.org/10.18326/register.v16i2.224-247>.

Nurhasan, Prahani, B. K., Suprpto, N., & Al Ardha, M. A. (2022). Artificial intelligence research during covid-19 pandemic: contributed to future education. *International Journal of Instruction*, 15(3), 229-248. <https://doi.org/10.29333/iji.2022.15313a>.

Ören, T., Turnitsa, C., Mittal, S. & Diallo, S.Y. (2017). Simulation-Based Learning and Education. In: Mittal, S., Durak, U., Ören, T. (eds) *Guide to Simulation-Based*

Disciplines. Simulation Foundations, Methods and Applications. *Springer, Cham*. https://doi.org/10.1007/978-3-319-61264-5_13.

Palavan, Ö. (2020). The effect of critical thinking education on the critical thinking skills and the critical thinking dispositions of preservice teachers. *Educational Research and Reviews*, 15(10), 606-627. <https://doi.org/10.5897/err2020.4035>.

Pnevmatikos, D., Christodoulou, P., Georgiadou, T. & Lithoxidou, A. (2023). Undergraduate Students' Conceptualization of Critical Thinking and Their Ideas for Critical Thinking Acquisition. *Educ. Sci.*, 13, 416. <https://doi.org/10.3390/educsci13040416>.

Podlasov, C. O. & Matviichuk, O. V. (2023). Application of ChatGPT in the teaching of physics to bachelor's students at a technical university. *Information Technologies and Learning Tools*, 97(5), 149-166. <https://doi.org/10.33407/itlt.v97i5.5374>.

Rudolph, J., Tan, S. & Tan, Sh. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1), 1-22. <https://doi.org/10.37074/jalt.2023.6.1.9>.

Rusandi, M., Ahman, Saripah, I., Khairun, D. & Mutmainnah. (2023). No worries with ChatGPT: building bridges between artificial intelligence and education with critical thinking soft skills. *Journal of Public Health*, fdad049. <https://doi.org/10.1093/pubmed/fdad049>.

Shiri, A. (2023). ChatGPT and academic integrity. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.4360052>.

Shulzhenko, Yu. M., Artemenko, Yu. O., Ponomarenko, O. O. & Kozhemiachenko, N. V. (2023). ChatGTP in the philologists training: practical use scenarios analysis and development of a catalog of hint templates. *Innovative Pedagogy*, 64, 35-43, 2023. <https://doi.org/10.32782/2663-6085/2023/64.2.7>.

Shyshkina, M. & Nosenko, Yu. (2023). Promising technologies with elements of AI for professional development of teaching. *Physical and Mathematical Education*, 38(1), 66-71. <https://doi.org/10.31110/2413-1571-2023-038-1-010>.

Sisilitsyn, Yu. O. & Osadchyi, V. V. (2023). Using ChatGPT in distance learning for beginners in programming. *Information Technologies and Learning Tools*, 97(5), 167-180. <https://doi.org/10.33407/itlt.v97i5.5277>.

Stokel-Walker C. & van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature*. Retrieved 25 February, 2023 from <https://www.nature.com/articles/d41586-023-00340-6>.

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays – should professors worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7>.

Susnjak, T. (2022). ChatGPT: The end of online exam integrity? *arXiv*. <https://doi.org/10.48550/arXiv.2212.09292>.

Taani, O. & Alabidi, S. (2024). ChatGPT in education: benefits and challenges of ChatGPT for mathematics and science teaching practices. *International Journal of Mathematical Education in Science and Technology*, 1–30. <https://doi.org/10.1080/0020739X.2024.2357341>.

The concept of artificial intelligence development in Ukraine, 2021. Retrieved 25 February, 2023 from <https://ips.ligazakon.net/document/KR201556?an=61>.

Theelen, H., van den Beemt, A. & den Brok, P. D. (2019). Classroom Simulations in Teacher Education to Support Preservice Teachers' Interpersonal Competence: A Systematic Literature Review. *Computers & Education*, 129, 14-26. <https://doi.org/10.1016/j.compedu.2018.10.015>.

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist D. & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, 11(3), 54. <https://doi.org/10.3390/jintelligence11030054>.

World Economic Forum jobs report. 10 Top Job Skills for 2025, 2020. Retrieved 25 February, 2023 from https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf.

Yandola, K. O. (2023). Using ChatGTP in the educational process. *Innovative Pedagogy*, 57(2), 261-265. <https://doi.org/10.32782/2663-6085/2023/57.2.53>.

Zhai, X. (2023). ChatGPT user experience: Implications for education. *Journal of Applied Learning and Teaching*, 6(1), 1-22. <https://doi.org/10.37074/jalt.2023.6.1.9>.