



Reasoning of generative artificial intelligence in solving mathematical logical problems

Tatjana Stanković ^{1, a, *}, Edin Liđan ^{2, b}

¹Belgrade Academy of Business and Arts, Belgrade, Serbia

²International Burch University, Sarajevo, Bosnia and Herzegovina

^atatjana.stankovic@bpa.edu.rs, ^bedin.lidan@ibu.edu.ba

Received: 1 October 2024; **Accepted:** 14 December 2024

Abstract

One of the tasks of mathematics education is to develop students' logical thinking. Logical problems that require a certain level of sharpness, ingenuity, and rational reasoning present a good challenge that can further motivate students to engage with mathematics. The progress of science and technology has led to the emergence of numerous tools that can be used in mathematics education. With the advent of artificial intelligence, these tools got a more active role as communication with users reached a new user-friendly dimension resembling human-to-human interaction. The unpredictability of the flow and the speed of development of artificial intelligence raises the question of whether it will ever reach or surpass the level of human sharpness and ingenuity. This paper illustrates inaccurate and partially accurate reasoning of artificial intelligence in solving logical problems.

Keywords: mathematical logical problems, generative artificial intelligence, ChatGPT, education.

MSC2020: 97U99, 97M10, 97M20

1. Introduction

The development of artificial intelligence has become intensively rapid and unpredictable. It raises the logical question of whether artificial intelligence will one day surpass human intelligence or not. Opinions are divided. Some papers state that machines will only be able to reach a part of the natural (human) way of thinking (Sokolowski, 1988), while others predict that artificial intelligence will reach the level of human intelligence by 2040 year and that it will become the so-called artificial superintelligence (by 2060) which will be greater than human intelligence (Owoc, Sawicka & Weichbroth, 2021). When it comes to mathematics, a serious challenge for artificial intelligence is represented by textual problems whose solutions require common sense reasoning (Davis, 2024). Human intuition and common sense reasoning can contribute to the better performance of artificial intelligence tools when it comes to solving mathematical problems, just as artificial intelligence tools can contribute to better performance, conclusions, and discoveries of mathematicians (Poola & Božić, 2023). Encouraged by the work (Poola & Božić, 2023), which indicates the importance of the connection between mathematical intuition and better performance of artificial intelligence using the example of ChatGPT, we came up with the idea of considering the behaviour of artificial intelligence when solving logical problems that require a certain amount

* Corresponding author

E-mail address: tatjana.stankovic@bpa.edu.rs

of ingenuity. For this purpose, we used ChatGPT version 3.5 due to its accessibility to the general public to solve 15 interesting problems.

2. AI and mathematical education

If we want to define artificial intelligence, we will come across many different definitions and approaches. In this paper, we have chosen the one that was given by The European Commission's High-Level Expert Group on Artificial Intelligence. This definition emphasizes the intelligent behaviour of the AI systems based on their environment analysis and partially autonomously acting in order to accomplish particular goals, and it distinguishes AI Systems in the virtual world that are entirely software-based and AI systems that are integrated into hardware devices ([The European Commission's, 2018](#)).

Lately, technology development has experienced its expansion and, therefore, a higher level of AI development. If we talk about using AI in science and, in general, in real-life situations, then we can talk about several different categories. General classification of AI can be done by its capability and its functionality. According to its capability, there are three types of AI: Artificial narrow intelligence, Artificial general intelligence and Artificial superintelligence. However, when we talk about its functionality, then we imply one of the following types: Reactive Machines (examples: IBM Deep Blue, Google AlphaGo, Netflix Recommendations Engine,...), Limited Memory (examples: Reinforcement learning, Long Short Term Memory (LSTMs), Evolutionary Generative Adversarial Networks (E-GAN),...), Theory of Mind and self-aware (examples: Self-driving cars, Customer service chatbots, Smart home devices, Industrial robotics). Each subtype of AI (whether it is based on its capability or its functionality) has its own subtype of AI. Generally, it could be said that AI can be classified into three groups: Artificial narrow AI (ANI), Artificial general AI (AGI) and Artificial super AI (ASI). In addition to all the previously mentioned major divisions of AI, we can also talk about Generative AI, Deep learning, Natural language processing (NLP), Weak vs. strong AI, Multimodal AI, Large language models (LLMs) ([GeeksforGeeks, 2024](#)).

We will be particularly interested in generative AI. Generative AI implies any form of tool that can create new content, which can be text, images, music and even some virtual world. One of the more general divisions of generative AI has four types: Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Autoregressive models and Transformers. Generative Adversarial Networks (GANs) are based on two neural networks-one of which creates new data based on existing data (Generator) and the other one that searches for similarity between new and existing data (Detector); the whole process of creating new data lasts as long as the Detector detects similarity among these two data sets. While Variational Autoencoders (VAEs) are based on a transformation of input data on coding and decoding in spaces of different dimensions, Autoregressive models create part by part of a data set by considering the context in which the previous part was created and predict the next part accordingly. Effectiveness in solving problems like language translation, text creation and image synthesis is provided by Transformers (whose main components are Encoder-Decoder Structure, Multi-Head Attention, Positional Encodings, Transformer Decoder, Position-wise Feedforward Networks in Decoder) ([GeeksforGeeks, 2024](#)).

From the very beginning of the development of mathematics, mathematicians and people who had to use and apply it were looking for tools that would help them in their daily life or in the process of learning or research. If we go back to the earliest beginnings of the development of mathematics, we will find papyrus, abacus, computers, and, to this day, modern technology that we call artificial intelligence. The application of tools is not exclusively a context related to mathematics as a science. Whether we are talking about education in general or we are talking about mathematical education, the tool as an aid has imposed itself as a rule and necessity. The use of artificial intelligence as a tool in education ([Guan, 2023](#)), and therefore in the teaching of mathematics, has proven to be effective in personalized learning and memorization of content in

which there is no need for reasoning or a logical path that leads us to a conclusion, which makes mathematics as a science particularly interesting in the context of AI research and development.

2.1. Generative AI in mathematical education in the case of ChatGPT

In this paper, we will specifically study the challenges of generative AI using the example of chatGPT and mathematical word problems. ChatGPT is a user-friendly generative AI, i.e., an NLP tool ([GeeksforGeeks, 2024](#)) that supports human-like communication among humans and AI in many different languages ([List of languages supported by ChatGPT, 2023](#)).

Although it is trained on a large amount of data, it keeps improving by exposing to more data so despite the limitations (reduced capability of meaningful communication, lack of comprehending distinction of human language and context of the conversation, absence of specific information that is outside the training set, absence or reduction of emotions i.e. of understanding emotional context) it has many features (in the human-like communication with user it can automatically provide quick and precise answers that are personalized and therefore customer service becomes better, the cost of additional hiring of people is reduced, thanks to its adaptability it can handle at the same time a large number of personalized conversations in different languages, etc.) so it impacts on many fields among which are academic, research and education ([Kalla et al., 2023](#)).

In education, ChatGPT is considered a tool that can improve the learning and teaching process. Thanks to its availability, accessibility, flexibility, and great knowledge base, it can be useful for individual learning by providing instant feedback. However, because of the lack of critical thinking and real-life experience, limited understanding of language and the context, potential bias and incorrect information, etc., this tool can also be unpredictable and may lead to misconceptions during the learning process. So, in order to maintain and raise the quality of education, the teaching and learning process using ChatGPT has to be under the teacher's supervision ([Oranga, 2024](#)). Especially in situations where ChatGPT can be used for unethical purposes, such as during the evaluation process ([Lo, 2023](#)).

During the process of using ChatGPT, inaccurate answers will pop up eventually. According to ([Butler, 2023](#)), the lack of ChatGPT that is reflected in providing incorrect answers can be a consequence of imprecise questions asked by the user. These questions (that are too short, unspecific, unclear enough, or without any context) will not provide enough information to ChatGPT to find the accurate answer in a large database ([Butler, 2023](#)). However, sometimes even when the question is precise, ChatGPT gives an inaccurate answer. This can be a consequence of incorrect information in the database set, misinterpretation of data (a statistically accurate conclusion that can be applicable in real-life), over-generalization ([Why ChatGPT Gives Wrong Answers: A Quick Guide, 2023](#)), etc. One of the interpretations of this situation considers inaccurate answers as a way of testing users by ChatGPT ([Montoro, 2023](#)). The interesting fact is that when it comes to accurate answers, ChatGPT will rather accept inaccurate answers as accurate than to stand up for the correct one ([The Ohio State University, 2023](#)).

ChatGPT is considered an innovative tool for teaching and learning mathematics ([Wardat et al., 2023](#)). In blended learning environments, ChatGPT requires the engagement of teachers in order to maintain ethical use and efficiency of the learning process since students highly believe that answers and explanations of the ChatGPT are correct and to avoid oversimplification ([Sánchez-Ruiz et al., 2023](#)). Although it may be useful in explanations, implementations, and evaluations, some teachers find it can lead to addiction and deceptions ([Uğraş et al., 2024](#)).

Although ChatGPT can successfully solve some problems, problem-solving strategies lack variety and flexibility ([Getenet, 2024](#)).

3. Mathematical problems and ChatGPT

Mathematics as a scientific discipline, school subject, or recreation is a synonym for something that is difficult for many people. In the context when we talk about mathematics as a school subject,

textual problems have a special place. Word problems are a big obstacle for many students, and they often fail to solve them. According to Hegarty and other authors (Hegarty et al., 1995), solving textual mathematical problems requires understanding the text that describes the mathematical problem, constructing a semantic structure, choosing numbers and keywords, constructing a problem model, developing a solution plan, and executing it. The question arises whether generative AI follows and offers elements that are proposed according to the given model and whether the application of AI will completely change the given model in the future. Numerous researchers have tried to find different models for solving mathematical word problems on smaller data sets (Mandal & Naskar, 2019), but besides the limited data set, challenges are also found in the non-procedural way of solving word problems based on excellent text comprehension (Pongsakdi et al., 2020) In this paper we will research how generative AI solves mathematical word problems on the example of using ChatGPT.

3.1. Methodology

In this paper, we investigate how generative AI can be implemented in the mathematical educational process in the context of learning and research. For this purpose, we carefully selected mathematical text problems that we wanted to solve using ChatGPT as one type of generative AI. The choice of researching the work of generative AI on the example of ChatGPT was not accidental; it was selected because ChatGPT is based on natural language processing (NLP) principles.

For this purpose, we have selected 15 math word problems that we gave chatGPT to try to solve. The description of the tasks and chatGPT's solution are given in the section Results. We would like to emphasize that for the purpose of this paper, all the problems and solutions were translated by ChatGPT with some authors' corrections.

3.2. Results

User [Problem 1.]: Four people played chess for 4 hours. How many hours did each participant play? (Zanimljivi zadaci, 2013)

ChatGPT Solution: Each participant played chess for 4 hours. Therefore, the total playing time is not divided among the participants; it is the time each participant spent playing. If four people played chess together for 4 hours, each played it for 4 hours.

User [Problem 2.]: Each stick has two ends. How many ends do one and a half sticks have? (Zanimljivi zadaci, 2013)

ChatGPT Solution: Each stick has two ends. So, if we have one and a half sticks, we can represent it as follows: 1 stick has 2 ends, 0.5 of a stick has 1 end (since it is half a stick, it has only one end). Therefore, 1 stick and 0.5 of a stick together have: $2 + 1 = 3$. One and a half sticks have a total of three ends.

User [Problem 3.]: A train with 10 carriages travelled 100 kilometres. How many kilometres did each carriage travel? (Zanimljivi zadaci, 2013)

ChatGPT Solution: Each carriage on the train travels the same distance as the train because all the carriages are connected and move together. So, if the train travelled 100 km, each of the 10 carriages also travelled 100 km.

User [Problem 4.]: A plank needs to be cut into six equal parts. How many cuts are required to achieve this? (Zanimljivi zadaci, 2013)

ChatGPT Solution: To cut the plank into six equal parts, you need to make five cuts. Each cut will create a new piece, so five cuts are required to get six parts. Here is how it works:

- After the first cut, you have 2 parts.
- After the second cut, you have 3 parts.
- After the third cut, you have 4 parts.
- After the fourth cut, you have 5 parts.
- After the fifth cut, you have 6 parts.

Therefore, five cuts are needed.

User [Problem 5.]: How can the number 66 be increased by its half without performing any mathematical operations? (Zanimljivi zadaci, 2013)

ChatGPT Solution: The number 66 can be increased by its half by writing it as “66 and a half.” So, $66 + 33 = 99$, but you can simply write it as 66.5, which means “66 and a half.”

Solution after a comment that the solution should be without performing any mathematical operations: The number 66 can be increased by its half without performing any mathematical operations by writing it consecutively: 66 becomes 666. When written as 666, it's as if you added half of 66 (33) in front of it, which gives 666 without performing any mathematical operations.

User [Problem 6.]: Is it correct to say 2 and 3 are 6, or 2 and 3 is 6? (Zanimljivi zadaci, 2013)

ChatGPT Solution: It is correct to say '2 and 3 are 5' or '2 plus 3 is 5.' However, if you use the expression '2 and 3,' it should be used in the context of addition, which would be '2 and 3 are 5.' From a grammatical perspective, if we consider 2 and 3 as two separate things, it would be correct to say '2 and 3 are 5' (though this is not mathematically accurate). For mathematical accuracy, it is correct to say '2 plus 3 is 5.'

User [Problem 7.]: A boy has as many brothers as he has sisters, and his sister has half as many sisters as she has brothers. How many brothers and sisters are there in this family? (Zanimljivi zadaci, 2013)

ChatGPT Solution: To solve this problem, let us set the following variables:

- B is the number of brothers.
- S is the number of sisters.

According to the problem, the boy has as many brothers as he has sisters. This means the boy has $B - 1$ brothers (since he does not count himself) and S sisters. On the other hand, the sister has half as many sisters as brothers. This means the sister has $S - 1$ sisters (since she does not count herself) and B brothers. So, according to the problem: $S - 1 = \frac{B}{2}$.

Now, we have two equations:

$$B - 1 = S \quad (1)$$

and

$$S - 1 = \frac{B}{2}. \quad (2)$$

Let us solve this system of equations. From equation (1):

$$S = B - 1.$$

Substitute S into the equation (2):

$$\begin{aligned} (B - 1) - 1 &= \frac{B}{2}, \\ B - 2 &= \frac{B}{2}. \end{aligned}$$

Multiply both sides by 2 to eliminate the fraction:

$$2(B - 2) = B, \quad 2B - 4 = B, \quad 2B - B = 4, \quad B = 4.$$

Now substitute B into the equation (2) to find S:

$$S = 4 - 1, \quad S = 3.$$

Therefore, the family has 4 boys (brothers) and 3 girls (sisters).

User [Problem 8.]: In a basket, there are 10 white, 7 red, and 5 green balls. How many balls must you take out, without looking, to ensure you have at least one ball of each color? (Logički zadaci, 2021).

ChatGPT Solution: To ensure that among the drawn balls, there is at least one of each color, we must consider the worst-case scenario. In the worst case, we could draw all the balls of one color, then all the balls of another color, before finally drawing a ball of the third color. We could first draw all 10 white balls. Then we could draw all 7 red balls. At this point, we have drawn 17 balls (10 white + 7 red) and still have not drawn a single green ball. Therefore, the next (18th) ball we

draw must be green, because all the white and red balls have already been drawn. So, the minimum number of balls that need to be drawn to ensure there are balls of all colors is:

$$10 + 7 + 1 = 18.$$

The answer is 18 balls.

User [Problem 9.]: If it is raining at midnight, can we expect sunny weather 72 hours later? (Logički zadaci, 2021).

ChatGPT Solution: Yes, it is possible to expect sunny weather after 72 hours. Weather conditions can change significantly over a 72-hour period. Even though it may be raining at a certain moment, the weather can change in the following days. Therefore, it is not unusual for sunny weather to follow a period of rain. However, the accuracy of weather forecasts depends on many factors and can vary in different situations.

User [Problem 10.]: Three gymnasts have the numbers 1, 3, and 6 on their shirts. What position should they take so that the resulting three-digit number is divisible by 7? (Petković, 1988, p. 2)

ChatGPT Solution: In order for the three-digit number formed by combining the numbers on the shirts to be divisible by 7, we can try all possible combinations of the numbers and check their divisibility by 7. We have three gymnasts with the numbers 1, 3, and 6. The combinations we can make are: 136, 163, 316, 361, 613, 631. Checking the divisibility of each combination by 7, we get: 136 is not divisible by 7, 163 is not divisible by 7, 316 is not divisible by 7, 361 is not divisible by 7, 613 is divisible by 7, 631 is not divisible by 7. Thus, the only combination that gives a three-digit number divisible by 7 is 613. The gymnasts should arrange themselves in the order of numbers 6, 1, and 3 (or reverse, as the sum remains the same) to form the number 613 on their shirts.

User [Problem 11.]: Three friends decided to go on a trip in a borrowed car. Since the fuel tank was empty, they pooled their money to buy gasoline. The first had enough money for 6 liters, the second for 8 liters, and the third for 10 liters. Just as they were leaving the gas station, a mutual friend came by and expressed his desire to join the trip. Upon returning from the trip, the fourth friend gave the three others 432 dinars as compensation for the gasoline. They divided the money proportionally to their contributions: the first took 108 dinars, the second 144, and the third 180 dinars. Is this division correct? (Petković, 1988, p. 12)

ChatGPT Solution: Let's check the correctness of this division by comparing the total amount the friends contributed to the gasoline with the total amount they received from the fourth friend after returning from the trip. The first friend contributed 6 liters, which we assume was at a price of x dinars per liter, so $6x$ dinars. The second friend contributed 8 liters at a price of x dinars per liter, so $8x$ dinars. The third friend contributed 10 liters at a price of x dinars per liter, so $10x$ dinars. In total, they contributed: $6x + 8x + 10x = 24x$ dinars. Now, let's look at how much money the fourth friend gave and then divide that amount according to the contributions of the first, second, and third friends: The fourth friend gave a total of 432 dinars. The amounts the first, second, and third friends received were $108 + 144 + 180 = 432$ dinars. Let's calculate the price per liter:

$$\frac{432}{24x} = \frac{18}{x}$$

dinars per liter. Thus, if the price per liter was 18 dinars, this division would be correct. However, without knowing the exact price per liter, we cannot conclude whether the division is correct or not. It is necessary to know the exact gasoline price they used for the calculation.

User [Problem 12.]: Girls are walking along a path. One is in front of two, one is behind two, and one is between two. How many girls are walking on that path? (Stojanović, 2012, p.11)

ChatGPT Solution: If one girl is walking in front of two, one is behind two, and one is between two, it means there are three girls on the path. Let's label them as A , B , and C . If A is in front of B , B is in front of C , and C is behind B , the arrangement on the path would look like this:

$$A - B - C \text{ or } C - B - A.$$

Therefore, there are three girls walking on the path.

User [Problem 13.]: Write 0 using three fours. (Logički zadaci, 2021).

ChatGPT Solution: 0 can be written using three fours as follows: $(4 - 4) \cdot (4 - 4) = 0$. In this expression, two fours are subtracted, resulting in 0. Then, the result (0) is multiplied by another $(4 - 4)$, which also gives 0.

User [Problem 14.]: We have two fuses whose lengths burn irregularly. Each of them burns completely in exactly one hour, regardless of which end is ignited. Can we use these fuses to measure a time interval of exactly 45 minutes? (Stojanović, 2012, p.25)

ChatGPT Solution: Yes, we can use these two fuses to measure exactly 45 minutes. Here's how:

1. Light one fuse at both ends and the second fuse at one end.
2. When the first fuse burns out completely, exactly 30 minutes have passed.
3. At that moment, the second fuse will have 30 minutes of burn time remaining.
4. Light the other end of the second fuse.
5. When the second fuse burns out completely, an additional 15 minutes will have passed.

In total, we have 30 minutes (from the first fuse burning from both ends) plus 15 minutes (from the second fuse burning from both ends), which equals exactly 45 minutes.

User [Problem 15.]: Four ships are sailing in such a way that each is exactly 1 kilometer away from each other. They are 1 cruiser, 1 battleship, and 1 aircraft carrier. The question is, what type is the fourth ship? (Maroti, 2010, p. 55)

ChatGPT Solution: Given that all the ships are exactly 1 kilometer apart from each other and have different characteristics, such as a cruiser, a battleship, and an aircraft carrier, the fourth ship should have a different characteristic not mentioned among the previously listed ships. One possible answer is that the fourth ship is a submarine, as the submarine was not explicitly mentioned as one of the previously listed types (cruiser, battleship, aircraft carrier). Therefore, the fourth ship could be a submarine.

3.3. Discussion and analysis of the results

The success of ChatGPT in solving problems is given by table (Table 1). ChatGPT got 9 correct solutions out of 15 and 6 incorrect out of 15, i.e. expressed as a percentage: 60% correct and 40% incorrect solutions.

The correct solutions illustrate accurate and well-explained answers given by ChatGPT. These explanations can be very useful to students during the learning process. Although the solution of Problem 15. is also accurate, its explanation is quite interesting. The way that ChatGPT reasoning works when solving this problem deviates from the expected mathematical reasoning. To be specific, ChatGPT does not recognize the traditional mathematical way of thinking, i.e. that the objects have to be at the vertices of the tetrahedron, and therefore, the solution is going to be a submarine. It approaches the problem from the language point of view (which is expected since it is an NLP tool), and it comes to an accurate answer by eliminating the mentioned objects from the list of naval vessels.

Table 1. Success of ChatGPT in solving problems.

Problem	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Correct solution	yes	no	yes	yes	no	yes	yes	yes	no	no	no	yes	no	yes	yes

Table 2. Mathematical word problems that chatGPT could not solve with types of mistakes.

Problems	Type of mistake
2. problem	misconception, applying proportionality with out considering real-life context
5. problem	considering number as a string, lack of experience in flipping number
9. problem	misconsidering real-life context
10. problem	in performing mathematical operation of division, lack of experience in flipping number
11. problem	misconception of real-life context
13. problem	misconception of the given conditions i.e. lack of comprehension of the given text

The incorrect solutions illustrate inaccurate answers that are based on misconception, misunderstanding of the context or even inaccurate basic operations (Table 2), so its explanations can be misleading and confusing to some students during the learning process. ChatGPT has knowledge about the concept of proportionality but it does not understand how and when to apply it in real-life situations (illustrated by ChatGPT Solution of Problem 2). The lack of human ingenuity has come to light in the Solution of Problem 5 (when ChatGPT comes up with a wrong solution, although in the previous step, it has calculated the right one). In spite of the fact that the Solution of Problem 10 also lies in flipping the number as in Problem 5, the mistake is based on incorrect division i.e. on the incorrect fact that the number 613 is divisible with 7. Insufficient comprehension of context and real-life situations and lack of common sense are also visible in the Solution of Problem 9 (it can't be concluded that since it is midnight, it will also be midnight and after 72 hours). Trying to find mathematical solutions to word problems that require common sense solutions (the Solution to Problem 11) or not considering all the given facts (the Solution to Problem 13) still exists.

Generalization of problems that require common sense, lack of understanding text and context of the given problem, lack of comprehending concepts in real-life situations and incorrect basic mathematical operations are still visible in some solutions given by ChatGPT

4. Conclusions and Implications for Future Research

This systematic ChatGPT as a tool can be suitable for the user if the user knows how to recognize illogicalities and inconsistencies in the solutions and answers to the problems or questions that are asked. When we study ChatGPT and its working style, we can approach it from many different aspects. The authors are based on different approaches to generative AI in education, but one of the interesting reviews of advantages and disadvantages is given by (Oranga, 2024). Our research indicates that ChatGPT has limited abilities to solve mathematical word problems as well as a lack of precision and reliability. Therefore, the user should approach carefully and with a critical attitude to its conclusions because they are not always correct. In other words, ChatGPT has a non-critical approach to problem-solving. In addition to the mentioned limitations, there is also a lack of experience in real-life situations that could help ChatGPT to understand the context of the given problem and there is a lack of understanding of the text itself i.e. requests of the given problem. But, it also has advantages that are reflected in detailed and step-by-step explanations, instant answers, human-like conversations with users, constant availability, etc.

According to some authors, ChatGPT makes mistakes for several reasons. Some of them are the result of inadequate use of ChatGPT (imprecise questions) and the others are the result of its design (incorrect information in the database set, misinterpretation of data, over-generalization etc.). ChatGPT certainly, along with its advantages, offers the possibility of eliminating the limitations that we have mentioned, and the path to it should also be sought not only in its design but also in the models that man has created, learned and used in his real-life.

References

- BUTLER, S. (2023, April 10). *6 Reasons ChatGPT is Giving You the Wrong Answers*. *How-To Geek*. <https://www.howtogeek.com/884058/reasons-chatgpt-is-giving-you-the-wrong-answers/>
- DAVIS, E. (2024). Mathematics, word problems, common sense, and artificial intelligence, *Bulletin of the American Mathematical Society*, 61 (2), 287–303. <https://doi.org/10.1090/bull/1828>
- GEEKSFORGEEKS. (2024, January 31). What is Generative AI? *GeeksforGeeks*. <https://www.geeksforgeeks.org/what-is-generative-ai/>
- GETENET, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: Implications for mathematics teaching in primary schools. *International Electronic Journal of Mathematics Education*, 19 (1), em0766, 1-12. <https://doi.org/10.29333/iejme/14141>

- GUAN, H. (2023). Advantages and Challenges of Using Artificial Intelligence in Primary and Secondary School Education, *Journal of Education, Humanities and Social Sciences: 2023 International Conference on Researches in Educational Technology and Psychological Sciences (RETPTS 2023)*, 22, 377-383. <https://doi.org/10.54097/ehss.v22i.12469>
- HEGARTY, M., MAYER, R. E., & MONK, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18–32. <https://doi.org/10.1037/0022-0663.87.1.18>
- KALLA, D., SMITH, N., SAMAAH, F., & KURAKU, S. (2023). Study and Analysis of Chat GPT and its Impact on Different Fields of Study. *International Journal of Innovative Science and Research Technology*, 8 (3), 827-833. Available at SSRN: <https://ssrn.com/abstract=4402499>
- List of languages supported by ChatGPT. (2024, November 22). <https://botpress.com/blog/list-of-languages-supported-by-chatgpt>
- LO, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Science*, 13 (4), 410. <https://doi.org/10.3390/educsci13040410>
- Logički zadaci (2021). *Scribd Web site*. <https://www.scribd.com/document/212717433/logicki-zadaci>
- MANDAL, S., NASKAR, S.K. (2019). Solving Arithmetic Mathematical Word Problems: A Review and Recent Advancements. In: Chandra, P., Giri, D., Li, F., Kar, S., Jana, D. (eds) *Information Technology and Applied Mathematics. Advances in Intelligent Systems and Computing*, 699. Springer, Singapore, 95-114. https://doi.org/10.1007/978-981-10-7590-2_7
- MAROTI, M. (2010). Mozgonetke: treniranje mozga uz razonodu. *Mladinska knjiga*, Beograd
- MONTORO, A. (2023, January 16). Is ChatGPT testing its users - Level Up Coding. *Medium*. <https://levelup.gitconnected.com/is-chatgpt-testing-its-users-643dff339f9f>
- ORANGA, J. (2023). Benefits of Artificial Intelligence (ChatGPT) in education and learning: Is ChatGPT helpful? *International Review of Practical Innovation, Technology And Green Energy (IRPITAGE)*, 3 (3), 46-50.
- OWOC, M. L., SAWICKA, A., & WEICHBROTH, P. (2021) Artificial Intelligence Technologies in Education: Benefits, Challenges and Strategies of Implementation, *Artificial Intelligence for Knowledge Management*, 37-58. https://doi.org/10.1007/978-3-030-85001-2_4
- PETKOVIĆ, M. (1988). Zanimljivi matematički problem. *Naučna knjiga*, Beograd.
- PONGSAKDI, N., KAJAMIES, A., VEERMANS, K., LERTOLA, K., VAURAS, M. & LEHTINEN, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM Mathematics Education* 52, 33–44. <https://doi.org/10.1007/s11858-019-01118-9>
- POOLA, I., & BOŽIĆ, V. (2023). Guiding AI with human intuition for solving mathematical problems in Chat GPT, *International Journal of Engineering & Scientific Research*, 11(7), 31-40.
- SÁNCHEZ-RUIZ, L. M., MOLL-LÓPEZ, S., NUÑEZ-PÉREZ, A., MORANO-FERNÁNDEZ, J. A. & VEGA-FLEITAS, E. (2023). ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. *Applied Sciences*, 13 (10), 6039. <https://doi.org/10.3390/app13106039>
- SOKOŁOWSKI, R. (1988). Natural and artificial intelligence, *Artificial Intelligence, Daedalus*, 117 (1), 45-64. Available at JSTOR: <https://www.jstor.org/stable/20025138>
- STOJANOVIĆ, V. (2012). Izoštri um. *Matematiskop*, Beograd.
- The European Commission's (2018). High-level expert group on Artificial intelligence a definition of AI: Main Capabilities and scientific disciplines. Brussels. Available at: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf
- The Ohio State University. (2023, December 7). ChatGPT often won't defend its answers – even when it is right. ChatGPT Often Won't Defend Its Answers – Even When It Is Right. <https://news.osu.edu/chatgpt-often-wont-defend-its-answers--even-when-it-is-right/>
- UĞRAŞ, H., UĞRAŞ, M., PAPADAKIS, S. & KALOGIANNAKIS, M. (2024). ChatGPT-Supported Education in Primary Schools: The Potential of ChatGPT for Sustainable Practices. *Sustainability*, 16 (22), 9855. <https://doi.org/10.3390/su16229855>
- WARDAT, Y., TASHTOUSH, M. A., ALALI, R. & JARRAH, M. A. (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>

Why ChatGPT Gives Wrong Answers: A Quick Guide - Jenni AI. (2023, October 10). <https://jenni.ai/chat-gpt/incorrect-answers-errors>

Zanimljivi zadaci. (2013, March 26). *Matematika*. <https://matematickizadaci.wordpress.com/zanimljivi-zadaci/>