

FOSTERING PROBLEM SOLVING AND CRITICAL THINKING IN MATHEMATICS THROUGH GENERATIVE ARTIFICIAL INTELLIGENCE

Alice Barana, Marina Marchisio and Fabio Roman

*Department of Molecular Biotechnology and Health Sciences, University of Turin
Via Nizza 52, 10126 Torino, Italy*

ABSTRACT

The spread of Artificial Intelligence (AI) has been recently generating worries among teachers and educators about the validity of assessment when students make use of AI tools to solve tasks. To tackle this issue, we propose mathematical problem solving activities to be carried out with the aid of ChatGPT, showing how problem solving and critical thinking continue to be pivotal in solving mathematical problems, even if this is performed with the aid of AI. After discussing theoretical frameworks on strategies of problem solving and phases of the critical thinking process, we present six problems of combinatorics that we submitted to ChatGPT. We also asked 40 university students to solve the six problems in group with the aid of ChatGPT during an international module on Problem Solving and Critical Thinking and collected the tutors' observations about the activities. Analyzing ChatGPT solutions and tutors' reflections, we show that the proposed activity requires problem solving and critical thinking to be accomplished. The results corroborate the idea that, instead of limiting the use of AI in education, it is possible to integrate it within learning and assessment to achieve the learning goals.

KEYWORDS

Artificial Intelligence, ChatGPT, Critical Thinking, Mathematics Education, Problem Solving

1. INTRODUCTION

The role of artificial intelligence (AI) in everyday life is increasingly extensive: nowadays, it helps us carry out fundamental tasks for the society, not only for specific jobs but also for general areas regarding the whole public, such as health and education (Lee, 2020). While on the one hand the advantages of AI are consolidated, on the other hand worries exist concerning the impact it could have once it becomes capable to substitute the human being to an important extent (Deranty and Corbin, 2022). An area in which this applies is education: how real is the risk that a student performs an assignment not on his/her own, but asking an AI to do it for them? (Crawford et al., 2023) The availability of a tool like ChatGPT, which gained a strong popularity during the last period of time also among the general public (Haleem et al., 2022), makes the question more topical than ever. Indeed, it has never been easier to ask the computer for a detailed text, or a full solution of a mathematical problem, starting from a simple query written in natural language, with no need of programming skills or other forms of specific interaction with the system. However, since this tool has spread, several cases of misinformation were documented (Farina and Lavazza, 2023), giving evidence that it is not generally possible to rely on its answers without thinking critically about these responses. In this paper, we will propose mathematical problem solving activities to be performed with the aid of ChatGPT. We will show how problem solving and critical thinking continue to be pivotal in solving mathematical problems, even if this is performed with the aid of AI, since the AI can often help the user up to a certain degree, but it is not so frequent that it returns a completely correct solution. Namely, in many instances, while the general setting of the procedure ChatGPT outputs is right, some key steps are incorrect, leading to errors that propagate through the proof. Proposing students to solve mathematical problems with the aid of this AI (or similar tools) can foster problem solving and critical thinking, for example while assessing the solution and recognizing where the AI fails, thus getting to the actual solution by correcting those steps. The structure of the paper is as follows: Section 2 outlines the theoretical framework within which this study is situated,

while Section 3 presents the research question and explores the methodology employed. Section 4 depicts the results, and Section 5 offers a thorough discussion. Finally, Section 6 concludes the paper with closing remarks.

2. THEORETICAL FRAMEWORK

2.1 Problem Solving and Critical Thinking

The role of problem solving in Mathematics is undisputed: especially when applications are involved, the development of a solution cannot usually be reduced to a simple sequence of actions to be applied mechanically. A comprehensive definition of problem solving involves the ability to understand the environment, identify complex problems, review related information to develop, evaluate strategies and implement solutions to build the desired outcome (Fissore et al., 2021). It is a consequence of this definition that problem solving does not only apply to Mathematics or professional life, but it is useful also for everyday life as citizens of the present world: the well-known 20th century philosopher and epistemologist Karl Popper once said *all life is problem solving* (Camiller and Popper, 1999). Some basic strategies of problem solving are (Wang and Chiew, 2010):

- Divide and conquer: subdivide a large complex problem into some smaller problems, where each of them can be solved more easily;
- Lateral thinking: approach solutions indirectly and creatively, without relying only on the rigid logic of step-by-step procedures;
- Analogy and reduction: starting from an analogous problem having a known solution, reconduct your situation to that problem;
- Bases of scientific method: assume a possible explanation in order to prove or disprove it (hypothesis testing), try to prove the unfeasibility of the problem in order to start from an absurd or more generally a failure (proof), test possible solutions until finding a right one (trial and error).

Critical thinking can be defined as the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action (Ennis, 2015). It can be considered constituted by the following phases (Changwong et al., 2018):

1. Describing: give a clear definition of what in consideration, by providing a proper specification of what was involved, where it happened and under which circumstances.
2. Reflecting: consider a topic under more than just one point of view, also by updating this consideration when new information or experience is available.
3. Analyzing: examine and explain the nature of something (e.g. a phenomenon), compare and contrast various elements, relate these evidences with topics.
4. Critiquing: detect strengths and weaknesses in arguments, under a neutral perspective.
5. Reasoning: use evidence and logic to prove or disprove an argument.
6. Evaluating: comment on what obtained in a wider sense, not just limiting to the reference subjects.

During the problem solving process, critical thinking has several roles: it permits to correctly look for data, it helps the choice of good strategies, and it allows arguing about the findings.

The use of problem solving for Mathematics with new technologies has been widely explored in the latest years (Barana and Marchisio, 2016; Marchisio et al., 2020; Marchisio et al., 2022a), being its benefits evident. With disruptive situations and tools, it becomes even more important to blend methodology and technology, in order to use the latter according to the goals devised with the former (Goldin and Katz, 2009; Marchisio et al., 2022b; Marchisio et al., 2022c).

2.2 Large Language Models and Generative AI Systems

AI tools such as ChatGPT belong to the category of generative AI systems, which in turn are based on large language models (LLMs). They are deep learning models capable of acquiring an extensive knowledge of a

language, by eliciting it from huge amounts of texts, mainly originating on the World Wide Web (Attardi, 2023). Starting from simpler tasks as predicting a word that concludes a sentence, they soon revealed to be capable of higher abilities, such as generating long texts from short instructions or solving scientific problems, by comparing an adequate set of possible answers. Collobert et al. (2011) presented a method of self-supervised learning aimed at representing the words' meaning, by providing to a neural network a sufficiently high number of sentences, from which developing the capability of recognizing patterns among the words constituting them. By representing every word with a long numerical vector, with each number standing for a particular nuance of meaning, it was possible to categorize words and to determine conceptual similarities, thanks to the comparison of these vectors. A limitation of that technique were the words possessing several and distinct meanings, whose meaning inside a sentence depends on the context; to overcome this hindrance, Vaswani et al. (2017) described an *attention mechanism* able to detect relations between words in a specific context. This is further applied by means of the so-called *Transformers*, which are models capable of preserving those relations, while producing an output starting from the user's input. The use of Transformers allowed, after just a few months, to process the natural language notably better than any previous technique, even those that required years of development and perfecting. This is possible because Transformers can be fine-tuned, by giving to a pre-trained model some new examples, allowing them to hastily adapt to the relative tasks. Regarding applications in solving mathematical problems, it is important to observe that LLMs show emergent abilities, appearing only when their dimension becomes particularly large, and consisting in advanced reasoning capabilities, such as the model being able to adapt a resolute path to a different setting (Wei et al., 2022). The use of generative AI now goes even beyond these capabilities, for instance by being able to also generate images starting from a brief description. This brings about ethics and security issues (Klenk, 2023), which are even more prominent when education is involved. However, the goal should be to properly integrate these tools by taking into account their limitations and merits, rather than turning the possible dangers into fears, thus demonizing them (Lim et al., 2023). Such features have been recently considered by several researchers (Ipek et al., 2023).

3. RESEARCH QUESTION AND METHODOLOGY

Our study is motivated by the following research question: *What strategies proper of problem solving and phases belonging to critical thinking do students activate when solving mathematical problems with the aid of ChatGPT?* In particular, we are interested in investigating how problem solving and critical thinking intervene when students interpret the outputs artificial intelligence gives and assess if they are actually correct or can anyhow help in some way. To answer the research question, we consider some mathematical problems in Combinatorics. They can be administered either to high school students, especially those possessing relevant interest in the sciences, or to university students, for example as preliminary tools for Probability and Statistics. These problems read as follows:

1. A regular polygon has n sides and $4n$ diagonals. Calculate n .
2. 23 officers participate to a military conference. If each participant shakes his or her hand to every other participant, how many handshakes will be overall performed?
3. Six squads are engaged in combat training. How many sessions are held, if in each session they confront each other two versus two, in every possible way?
4. n servicemen are involved in a muscle strengthening program, organized such that every day 4 of them are present, and after 13 days everyone has been present exactly once with each other. What is the value of n ?
5. In a platoon composed of 36 soldiers, 19 of them are going to have field training during the next week (in the day), while 18 of them are going to take part in a workshop (in the evening). Six of them will be on leave, so they are not going to have field training nor take part in the workshop. How many soldiers are going to both have field training and take part in the workshop?
6. A test is made of 15 multiple-choice questions, admitting as answers three alternatives, let them be A, B, C. For the sake of unpredictability, it is required that at least one correct answer is A, at least one is B, and at least one is C. In how many ways is it possible to construct the (ordered) list of correct answers?

The use of elementary combinatorial techniques is required to solve these problems, but their solutions are not immediate, being them the outcome of a multistep procedure, rather than a single computation.

First, we repeatedly asked ChatGPT (GPT-3.5 architecture) to solve the problems and we analyzed what it gave as outputs, in terms of correctness, consistency and clarity. In particular, we are interested in how students can detect possible errors (which are not unlikely to occur) by thinking critically and applying problem solving strategies, with the goal of devising a correct solution after the human intervention, but starting from the AI output. As an important note, it should be made clear that differently from other studies, our goal is not to assess the success rate of ChatGPT, or to perform a statistical inquiry; instead, we aim at showing some possibilities that can occur, and how students can interpret them according to their aforementioned objective. Moreover, we experimented the six problems with international university students enrolled in degree courses in strategic sciences during an international module on Problem Solving and Critical Thinking. Students were asked to solve the problems in groups using ChatGPT and discuss the solutions. Four tutors facilitated the activities and filled a questionnaire at the end, constituted of open questions aimed at capturing their insights on how problem solving and critical thinking were activated during the activities. In particular, we analyzed the tutors' answers to the questions "Which Strategies of Problem Solving did they adopt?" and "Which Phases of critical thinking did they perform?", selecting references to the use of AI in these processes, in order to confirm the preliminary results.

4. RESULTS AND DISCUSSION

4.1 Solving the Problems with ChatGPT

We started by repeatedly submitting Problem 1 to ChatGPT. The pivotal step is that a polygon, not necessarily regular, having n sides, has $n(n-3)/2$ diagonals, and so it reduces to the equation $n(n-3)/2 = 4n$. Most of the times, the AI correctly performed this step, then arriving by elementary algebraic steps to the solutions $n = 0$ and $n = 11$, of which ChatGPT explicitly states that only the latter is feasible, since a polygon cannot have zero sides (see Figure 1). Unfortunately, sometimes it commits an error by writing that a n -sided polygon has $n(n-3)$ diagonals, thus writing a different equation resulting in the wrong solution $n = 7$. Furthermore, the procedure ChatGPT outputs states that the formula holds "for any regular polygon", while in fact this is not limited to them, being rather valid at least for any *convex* polygon.

By performing the same with Problem 2, again most of the times the solution was correct: two officers out of a group of 23 can be chosen in $(23 \cdot 22)/2 = 253$ ways, so there will be 253 handshakes. However, in some instances it reasoned differently: the first officer will shake hands with the remaining 22 officers, the second officer will shake hands with the remaining 21 officers (excluding the first officer), the third officer will shake hands with the remaining 20 officers (excluding the first and second officers), and so on. This led to the sum $22 + 21 + \dots + 2 + 1$, which yields again 253, but in an instance ChatGPT computed it wrongly, as Figure 2 depicts. Indeed, the AI wrote the formula for summing that arithmetic series as $(23/2) \cdot (22+1)$ instead of $(22/2) \cdot (22+1)$: the terms are 22 and not 23, or alternatively the 23^{rd} is 0, so it would have been $(23/2) \cdot (22+0)$.

Regarding Problem 3, on the contrary ChatGPT kept giving the incorrect solution 15, as choosing two squads out of a group of six for each session: $(6 \cdot 5)/2 = 15$. In fact, the squads to be chosen are four and not two, since the confrontations are two versus two. Moreover, for each choice fixing four squads, let us say A,B,C,D, we can let them confront as A,B vs C,D, or A,C vs B,D, or A,D vs B,C, so there are three ways. The total number of sessions is thus $3 \cdot (6 \cdot 5 \cdot 4 \cdot 3)/4! = 45$, but the AI was unable to output this answer.

Concerning Problem 4, ChatGPT provided the right solution $n = 13$ every time except the first one, when it resorted to $n \cdot (n-1)/2 = 4$ (having positive solution $n = 4$), instead of $n \cdot (n-1)/2 = 13 \cdot (4 \cdot 3)/2$, thus ignoring the number of days and the coupling of the servicemen present every day.

For Problem 5, $36 - 6 = 30$ soldiers will be not on leave, so if respectively 19 and 18 are engaged in the two activities, then $19 + 18 - 30 = 7$ are engaged in both. ChatGPT was unsuccessful on this: an error it committed during a try was to ignore the soldiers on leave (thus answering $19 + 18 - 36 = 1$), even after a first part of reasoning in which it considered them. Another error in a subsequent try regarded the confusion of quantities, first by stating that $19 + 18 - 6 = 31$ soldiers are engaged in at least one activity, while they are

actually the 30 not on leave, and then by considering them as soldiers with no participation, finally answering a wrong $36 - 31 = 5$. In one case, it ignored both the soldiers on leave and their total number, by simply affirming that at most $\min(18,19) = 18$ soldiers could be doubly engaged, using only information from those numbers.

Finally, in Problem 6, ChatGPT gave an almost correct solution, by just committing the error of considering a single case in which A, B and C are all missing, which cannot exist: the number of valid combinations is thus given as $3^{15} - 3 \cdot 2^{15} + 3 - 1$, where the final -1 is actually wrong. Nevertheless, the AI failed to compute the result as a number, by stating that it equals to 14,348, when in fact it is equal to 14,250,605 (14,250,606 is the solution to the problem, by not considering -1). The fact that ChatGPT was able to perform almost correctly a task which had been traditionally deemed as difficult for an automated system, such as solving an articulated mathematical problem (Problem 6 was the most difficult of the list for a human solver), but failed in a simple algebraic computation, is noteworthy. Indeed, it is representative of the nature of this generative AI system, which possesses more data retrieval and assemblage potential, rather than computational capabilities.

F A regular polygon has n sides and $4n$ diagonals. Calculate n .

Let's start by understanding the formula for calculating the number of diagonals in a regular polygon.

For any regular polygon with n sides, the number of diagonals can be calculated using the formula:

$$\text{diagonals} = n * (n - 3) / 2$$

Given that the number of diagonals is $4n$, we can set up the equation:

$$4n = n * (n - 3) / 2$$

To simplify the equation, let's multiply both sides by 2:

$$8n = n * (n - 3)$$

Expanding the right side:

$$8n = n^2 - 3n$$

Rearranging the equation:

$$n^2 - 11n = 0$$

Factoring out n :

$$n(n - 11) = 0$$

This equation has two solutions: $n = 0$ and $n = 11$. However, a regular polygon cannot have 0 sides, so we discard that solution.

Therefore, the regular polygon has $n = 11$ sides.

Figure 1. The correct solution of Problem 1 as provided by ChatGPT

4.2 Analysis of the Problem Solving and Critical Thinking Activities

The solutions ChatGPT gave as outputs are first discussed in relation with the problem solving strategies we presented in the theoretical framework. For Problem 1, a student shall split how to determine the equation by means of the formula associating the number of diagonals a polygon possesses to its number of sides, from how to solve the equation itself. This fits into the “divide and conquer” strategy of problem solving. Moreover, to ascertain the incorrectness of the formula if returned wrong, students can use a “base of the scientific method”: finding a counterexample. Indeed, a student should be able to easily recall that

quadrilaterals possess two diagonals, since they are widely used during geometry classes at compulsory schooling. By substituting $n = 4$ in the expression $n(n-3)$, the result is 4 rather than 2, thus directly suggesting that something is wrong. On the other hand, “lateral thinking” could help in assessing the validity of the formula if returned correct: if a student does not know or does not remember how to proceed with the combinatorial theory, s/he can rely on a bit of creativity, for instance by trying some cases with a low number of sides (of which $n = 4$ can be one), and then finding a motivation for which the expression $n(n-3)/2$ is valid for every n . Finally, to prove that the formula holds also for polygons which are not regular, “analogies” can be used: for example, does the number of diagonals change if a vertex is dragged, a modification which keeps intact the number of sides?

F 23 officers participate to a military conference. If each participant shakes his or her hand to every other participant, how many handshakes will be overall performed?

 To calculate the number of handshakes when each participant shakes hands with every other participant, we can use a simple formula.

In this case, there are 23 officers. The first officer will shake hands with the remaining 22 officers. The second officer will shake hands with the remaining 21 officers (excluding the first officer). The third officer will shake hands with the remaining 20 officers (excluding the first and second officers), and so on.

The total number of handshakes can be calculated by summing the number of handshakes for each officer.

$$\text{Total handshakes} = 22 + 21 + 20 + \dots + 2 + 1$$

This is an arithmetic series with a common difference of 1 and a first term of 22. We can use the formula for the sum of an arithmetic series:

$$\text{Total handshakes} = (n/2) * (\text{first term} + \text{last term})$$

where n is the number of terms.

In this case, $n = 23$ (since there are 23 officers), the first term is 22, and the last term is 1.

$$\begin{aligned} \text{Total handshakes} &= (23/2) * (22 + 1) \\ &= 23/2 * 23 \\ &= 23 * 23/2 \\ &= 23 * 11.5 \\ &= 264.5 \end{aligned}$$

Therefore, there will be a total of 264.5 handshakes performed overall at the military conference. Note that since a handshake is a physical interaction between two individuals, the number of handshakes must be a whole number. Therefore, we round down the result.

Rounded down, the total number of handshakes will be 264.

Figure 2. ChatGPT solving Problem 2: right procedure, wrong computation

In Problem 2, it is again important splitting how to write a formula from how to compute it (“divide and conquer”), especially if ChatGPT returns a sum rather than what follows from a direct combinatorial approach. In this case, the “lateral thinking” strategy can intervene in finding the possible incorrectness of the solution. Indeed, an inherent contradiction is present where a sum of integers yields the formula $(23/2) \cdot (22+1)$ which, if computed explicitly, equals to $23^2/2 = 529/2 = 264.5$ (see Figure 1): summing integer numbers cannot result in obtaining a non-integer number, a fact that students definitely know from elementary arithmetic, and they just have to recall it in a context probably different from where they learned

it. Students should also note that the final rounding down to make the number feasible in the context (yielding 264) is just an attempt to forcefully “rectify” the result in order to obtain an integer, which does not remove the error the aforementioned fact implies. Note also that this can be related with the search for counterexamples allowing to deem the wrong formula as incorrect in Problem 1.

Concerning Problem 3, it is useful to see how ChatGPT approaches the problem, but then students are a bit more on their own than in Problems 1 and 2. Indeed, since the AI keeps choosing two squads rather than four, the student can perform an “analogy” and use the combinatorial formula proposed, but with 4 instead of 2 where appropriate. Incidentally, here $(6 \cdot 5)/2 = (6 \cdot 5 \cdot 4 \cdot 3)/4! = 15$, but this is not the point: conceptually, working with 2 is wrong. Moreover, the problem does not end with this computation, given the three ways in which the four squads can be paired two versus two: students thus have to divide the part in which the four squads are *chosen* from the part in which they are *paired* (with the choice performed).

About Problem 5, the presence of contradictions is clear seeing how ChatGPT changed premises on which reasoning during the procedure, giving proof of inconsistency (that pertains to the “scientific method”). Somewhat similarly to Problem 3, some sort of “analogy” could be performed by starting from the formula summing the numbers relative to the two engagements and then subtracting a third number: indeed, the AI suggests it as the correct tool to obtain a solution, its errors lying in the inability to correctly write the number after the minus sign. Analogous reasoning holds for Problems 4 and 6.

Now, we can discuss the solutions with regard to the critical thinking phases. The first three phases are implied by the ways in which ChatGPT approaches the problems, since they foster the student to devise an outline of the situation (“describing”), to evaluate its implications (“reflecting”) and to assess its adequateness with respect to the objectives (“analyzing”). The “critiquing” phase can emerge well in situations such as the presence of intermediate steps being somewhat different from what expected, for example in the steps leading ChatGPT to give the wrong answer to Problem 4: the absence of the days’ number in the formula $n \cdot (n-1)/2 = 4$ is not necessarily a contradiction, but it should bring up some suspect, thus belonging to the assessment of the likelihood of rationales to be solid and consistent. On this line, the “reasoning” phase is prominently clear where contradictions are instead actual, like in the errors occurring during an instance of solving Problem 2, when the number 264.5 can be definitely disproved as possibly being the sum of integers. Finally, the “evaluating” phase does not directly descend from what ChatGPT gives as outputs, since it limits itself to provide the procedures allowing to solve the problems, but nonetheless students can consider them as a starting point to further discuss the solutions.

4.3 Analysis of the Tutors’ Questionnaire

Concerning problem solving, Tutor A wrote “I suggested them to use ChatGPT when they were in difficulty, and then try to understand and evaluate its solution”, implying that ChatGPT supported students in the choice of the solving strategy and in lateral thinking when they did not have any idea on how to approach it. Starting from the AI suggestion, students had to use “bases of scientific method” to prove or disprove it. The same tutor also adds: “Sometimes they divided problems into smaller parts and used ChatGPT to better understand some concepts”, making clear how the tool can be used to support the “divide and conquer” solving strategy. Tutor B noted that “2 students used ChatGPT to solve problems, for example to do some computations, and checked that the answers provided by their colleagues (who used a traditional approach to solve problems) were correct.” This answer shows how ChatGPT was used to solve steps of the solving process (supporting the “divide and conquer” strategy) and in the “bases of scientific method”, testing hypothesis. Tutor C did not mention ChatGPT in the problem solving strategies. Tutor D wrote “the use of ChatGPT fostered in them “analogy”, since starting from the solutions the AI posted, being often not completely correct, they were able to devise the right ways by taking AI solutions’ paths as a reference, correcting them thanks to the knowledge they had”.

Concerning critical thinking, tutor A declared to have observed the following phases: “Analyzing and critiquing when students examined the solutions provided by ChatGPT to analyze coherence and correctness; reasoning when discussing about the solutions, in particular when they used different strategies to solve problems; the evaluating phase was supported by ChatGPT answers since they helped students to interpret the results, thus moving from reasoning to opinions”. Tutor B did not mention ChatGPT when describing the students’ critical thinking process. Tutor C wrote “They subdivided some sub-tasks to evaluate critically the single step. It is mainly relevant to note that there was (like I suggested) double check on calculations also

using ChatGPT, because they made several mistakes at the beginning and they were asking me about the correct solution, but this is something that does not happen in real scenarios, so they had to check and verify on their own the correctness of results”. The use of AI to check results, which Tutor B mentioned in the problem solving strategies, is here recalled to support the “reflecting”, “analyzing” and “critiquing” phases of critical thinking. Tutor D noticed that “the “examine” phase intervened before trying analogies while assessing whether the solutions the AI gave were correct or not (and in case not, where they failed), and so on”. Moreover, he added that “they used also ChatGPT, not only to get ideas, but also to compare their reasoning with the solutions the AI provided”.

Summing it up, students mainly used ChatGPT to support problem solving in finding ways to solve problems and testing solutions; ChatGPT supported the critical thinking process, in particular when they checked the correctness of the solutions proposed by the AI or, vice versa, when using the tool to check their own solutions. This yields scientific and practical implications, starting from a consideration the collected results allowed us to infer: the AI did not act as a tool to *substitute* the human, but rather it *flanked* the students, without undermining their role as an active part of the processes. Consequently, there was no harm in letting students to deal with ChatGPT, while they had in practice the possibility to complement their cogitating, and the relative benefits could be scientifically studied. According to the tutors’ responses, not every student took advantage of the AI: this might represent a current limitation, in the sense that some students could not have perceived as helpful the aid ChatGPT provided. However, these tools are still a novelty, and there is yet plenty of opportunities to instill in people confidence in them.

5. CONCLUSION AND FUTURE WORK

As a possible answer to the alarm that has recently spread among teachers due to the availability of AI that could hinder the validity of assessment, in this paper we proposed to ask students to use ChatGPT in mathematical problem-solving activities. In particular, we show that this AI can support the problem solving strategies and the critical thinking process. We answered to the research question “*What strategies proper of problem solving and phases belonging to critical thinking do students activate when solving mathematical problems with the aid of ChatGPT?*” by analyzing the solutions that ChatGPT provides to six problems of combinatorics in the perspective of problem solving and critical thinking. Moreover, we analyzed the observations that tutors made after experimenting the same six problems with university students during an international module. The results show that all strategies of problem solving and all phases of critical thinking can be affected by this kind of activity, since students have to check the correctness of ChatGPT solutions, not always correct, or check their own solutions through ChatGPT. Moreover, the activity also helped them find new and different approaches to solve problems. In line with Crawford et al. (2023) and Lim et al. (2023), we are convinced that, instead of limiting the use of AI for learning and assessment, it is possible to integrate it during the activities and find new ways to evaluate the achievement of learning goals, thus turning the threat into an ally. This also implies that the misinformation risks, which we saw as concrete while doing Mathematics like they are in other contexts as highlighted by Farina and Lavazza (2023), can be handled constructively, by taking advantage from what is right, and at the same time recognizing and correcting what is wrong. The advanced reason capabilities Wei et al. (2022) described, which manifested also here, indicate that the “right” portion is generally extended enough to be a safe starting point for students to positively use the outputs provided by the AI.

This research work could be further developed, for example by videorecording the students when dealing with this kind of activity, in order to observe with high precision the kind of support that ChatGPT gives to the problem solving and critical thinking processes. In addition, other generative AI systems such as You.com, Bing AI and Bard are available as ChatGPT is, so a comparison between them can be set up. Moreover, the activities could be extended to different kinds of problems and different areas of Mathematics.

ACKNOWLEDGEMENT

The authors would like to thank the bank foundation Compagnia di San Paolo, which financially supports many of the initiatives at the University of Torino, in particular the project OPERA Open Program for Educational Resources and Activities, in which this research took place.

REFERENCES

- Attardi, G., 2023. Il Bello, il Brutto e il Cattivo dei LLM. *Mondo Digitale*, 22(101).
- Barana, A. and Marchisio, M., 2016. From digital mate training experience to alternating school work activities. *Mondo Digitale*, 15(64), pp. 63-82.
- Camiller, P. and Popper, K., 1999. *All Life is Problem Solving*. Routledge, London, DOI: 10.4324/9780203431900
- Changwong, K. et al, 2018. Critical thinking skill development: Analysis of a new learning management model for Thai high schools. *Journal of International Studies*, 11(2), pp. 37-48, DOI: 10.14254/2071-8330.2018/11-2/3
- Collobert, R. et al, 2011. Natural Language Processing (Almost) from Scratch. *The Journal of Machine Learning Research*, 12, pp. 2493-2537, DOI: 10.5555/1953048.2078186
- Crawford, J., Cowling, M., & Allen, K., 2023. Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice*, Vol. 20, no 3, pp. 1-19. DOI: 10.53761/1.20.3.02
- Deranty, JP. and Corbin, T., 2022. Artificial intelligence and work: a critical review of recent research from the social sciences. *AI & Soc.*, DOI: 10.1007/s00146-022-01496-x
- Ennis, R.H., 2015. Critical thinking: a streamlined conception. In: Davies M., Barnett R. (eds.) *The Palgrave Handbook of Critical Thinking in Higher Education*, pp. 31-47. Palgrave Macmillan, New York, DOI: 10.1057/9781137378057
- Farina, M. and Lavazza, A., 2023. ChatGPT in society: emerging issues. *Frontiers in Artificial Intelligence*, vol. 6, DOI: 10.3389/frai.2023.1130913
- Fissore, C. et al, 2021. Development of Problem Solving Skills with Maple in Higher Education. In: Corless R.M., Gerhard J., Kotsireas I.S. (eds.) *Maple in Mathematics Education and Research. MC 2020. Communications in Computer and Information Science*, vol. 1414, pp. 219-233. Springer, Cham. DOI: 10.1007/978-3-030-81698-8_15
- Goldin, C.D. and Katz, L.F., 2009. *The Race between Education and Technology*. Harvard University Press, Cambridge, UK, DOI: 10.2307/j.ctvjf9x5x
- Haleem, A. et al, 2022. An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2(4), 100089, DOI: 10.1016/j.tbench.2023.100089
- Ipek, Z.H. et al, 2023. Educational Applications of the ChatGPT AI System: A Systematic Review Research. *Educational Process*, 12(3), pp. 26-55, DOI: 10.22521/edupij.2023.123.2
- Klenk, M. Ethics of Generative AI and Manipulation: A Design-Oriented Research Agenda. *Social Science Research Network*, DOI: 10.2139/ssrn.4478397
- Lee, R., 2020. *Artificial Intelligence in Daily Life*. Springer, Singapore, DOI: 10.1007/978-981-15-7695-9
- Lim et al, 2023. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790, DOI: 10.1016/j.ijme.2023.100790
- Marchisio, M. et al, 2020. Teaching Mathematics in Scientific Bachelor Degrees Using a Blended Approach. *Proceedings of IEEE 44th COMPSAC Conference*, pp. 190-195. DOI: 10.1109/COMPSAC48688.2020.00034
- Marchisio, M. et al, 2022a. Teaching Mathematics to Non-Mathematics Majors through Problem Solving and New Technologies. *Education Sciences*, 12(1):34. DOI: 10.3390/educsci12010034
- Marchisio, M. et al, 2022b. Teachers' perception of higher education in a transition scenario. *Proceedings of IEEE 46th COMPSAC Conference*, pp. 139-144. DOI: 10.1109/COMPSAC54236.2022.00028
- Marchisio, M. et al, 2022c. Teachers' digital competences before and during the COVID-19 pandemic for the improvement of security and defence higher education. *16th International Conference on e-Learning (EL2022) – Held at the 16th Multi-Conference on Computer Science and Information Systems (MCCSIS2022)*, pp. 68-75.
- Vaswani, A., et al, 2017. Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000-6010. DOI: 10.5555/3295222.3295349
- Wang, Y. and Chiew, V., 2010. On the cognitive process of human problem solving. *Cognitive Systems Research*, 11(1), pp. 81-92, DOI: 10.1016/j.cogsys.2008.08.003
- Wei, J., et al, 2022. Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*. DOI: 10.48550/arXiv.2206.07682