ChatGPT's Capability in Solving Mathematical Modelling Problems

Nemanja Vučićević[1][0000-0002-4903-7280], Aleksandar Milenković[1]* [0000-0001-6699-8772], and Marina Svičević[1][0000-0003-2791-3849]

[1] University of Kragujevac, Faculty of Science, Radoja Domanovića 12, 34000 Kragujevac, Serbia
nemanja.vucicevic@pmf.kg.ac.rs,
aleksandar.milenkovic@pmf.kg.ac.rs, marina.svicevic@pmf.kg.ac.rs

Abstract. In this paper, we examine the capability of ChatGPT in solving problems based on mathematical modelling. In line with the research objective, we assigned the free version of ChatGPT (GPT-4o) six problems that fourth-year mathematics students, majoring in Applied Mathematics, had solved as part of their exam. We analyzed whether the following steps were properly carried out: constructing a mathematical model, the problem-solving process in the mathematical context, interpreting the solution in relation to the problem, and finally, validation. The results were then compared with the work of the aforementioned students. The findings indicate that, in all parts of the solution and for all tasks, the free version of ChatGPT was either completely or partially successful. Nevertheless, it achieved results that were weaker than those of the students. Based on these findings, we can say that solutions to problems addressed through mathematical modelling – which, in the second step, involve solving first-order differential equations, game theory problems, and linear programming tasks – can be used to some extent as student support. However, it is equally necessary to work on the specialization of LLMs that would solve problems of this level of complexity in accordance with the mathematical modelling process.

Keywords: Mathematical Modelling, ChatGPT, Student Performance Comparison

1 Introduction

In recent times, an increasing number of researchers in the fields of education, engineering, and computer science have been exploring the potential and capabilities of large language models (LLMs) in solving various tasks through AI tools. Contemporary research mainly investigates the extent to which specific tools can successfully address different challenges and solve concrete tasks and problems and often compares the performance of different models in executing specific user instructions. Among these tools, ChatGPT is probably the most widely used.

Mathematical modelling is gaining increasing prominence in mathematics education. On one hand, it is important to introduce future mathematicians – particularly those seeking careers in industry – to the principles of mathematical modelling so that

* Corresponding authors' e-mail: aleksandar.milenkovic@pmf.kg.ac.rs

Research Paper
DOI: https://doi.org/10.46793/eLearning2025.184V
Part of ISBN: 978-86-89755-37-4

they can successfully translate real-world problems from various fields of science and industry into mathematical contexts, solve them, and interpret and discuss the results within the original context. Likewise, prospective mathematics teachers should acquire the same competencies and develop the pedagogical skills needed to teach their students the step-by-step process that characterizes mathematical modelling.

This raises the question of to what extent ChatGPT can effectively solve mathematical modelling problems, to serving as an additional resource in the education of future mathematicians and, in particular, as support in solving problems from various domains through the application of mathematical tools. This study aims to contribute to this emerging area of research.

2 Literature Review

One of the widely accepted processes or cycles of mathematical modelling was proposed by Blum and Leiß [1]. The cycle they suggested consists of six phases and six transitions between these phases. According to their methodology, the first step is to understand the real-world problem and construct the so-called situation model (1). This situation model is then simplified – specifically, the problem is structured, and the necessary assumptions are formulated (2). The next step, which is particularly important in this cycle, involves constructing a mathematical model through a process of mathematization (3). In this step, the process shifts into a mathematical context. The mathematical problem thus formed is then solved, yielding the required mathematical results (4). These mathematical results are then considered and examined in the context of the real-world problem to determine how they can be effectively used to solve the initial problem; through interpretation, results corresponding to the initial real situation are obtained (5). Finally, the results obtained should be validated (6). Within this validation process, the modelling steps should be reconsidered by re-examining the assumptions, the mathematical models, their solutions, and the correctness of the calculations. Although it represents an important aspect of learning, the meta-analysis conducted by Schukajlow et al. [2] found that only 3% of research papers published in leading journals on mathematics education over a six-year period focused on the topic of mathematical modelling. Furthermore, most of these papers were empirical studies with a qualitative research design. Sen Zeytun et al. [3] conducted a qualitative study involving six mathematics students in order to identify the sources of students' difficulties in learning mathematical modelling. Based on their results, it can be concluded that the factors contributing to students' difficulties can be divided into contextual and individual categories. Contextual factors refer to the students' lack of experience in solving problems through mathematical modelling, while the individual factors relate to the time constraints for completing the tasks. Individual factors also include gaps in mathematical knowledge and skills, difficulties in establishing connections between the real-world and mathematical contexts (and vice versa), students' tendency to focus solely on the result, as well as their inability to adequately organize their work [3].

ChatGPT is a Natural Language Processing (NLP) model developed by OpenAI that uses a large dataset to generate responses to students' questions, prompts, and feedback

[4]. There are many studies that explore the capabilities of ChatGPT and other OpenAI models in solving mathematical tasks and problems. For example, OpenAI models have been tested in solving complex mathematical competition problems. Interestingly, when solving problems of lower complexity but with a higher degree of non-standard structure, OpenAI tools perform at the level of average competition participants [5]. However, when tackling more demanding mathematical problems intended for high school students, the o1 and o3-mini models would win awards when competing among the best high school students [6].

ChatGPT can solve problems in the field of linear algebra, compute integrals, and solve differential equations when given appropriate instructions [7]. When it comes to differential equations, it was found that ChatGPT correctly solved 72% of the given first-order differential equations, made minor errors in 20% of them, and in the remaining 8% it made serious computational or methodological errors. In solving second-order differential equations, it solved 64% of the problems completely correctly, made minor errors in 24% of them, and made a serious computational or methodological error in every eighth problem [8]. Zhu et al. [9] explored ChatGPT's compatibility in decision-making when solving game theory problems, which are also widely used in mathematical modelling.

Some research on integrating LLMs into solving mathematical modelling tasks has already been conducted. For instance, Huang et al. [10] introduced a LLM model called Mamo, which has integrated solvers aimed at validating solutions and evaluating the correctness of the entire mathematical modelling process.

3 Research question

Considering the importance of mathematical modelling for the education of mathematicians – regardless of whether their future profession will be in industry or in teaching – as well as the potential of ChatGPT for solving tasks of various levels in terms of mathematical content and complexity, we aimed to examine the accuracy and overall problem-solving process of tasks based on mathematical modelling generated by ChatGPT. This gave rise to the following research questions.

1. To what extent does ChatGPT successfully solve mathematical modelling problems related to the application of differential equations, game theory and linear programming?
2. Are there differences in the performance between mathematics students who attended the Mathematical Modelling course and ChatGPT in terms of accuracy and the structure of solutions within the context of the mathematical modelling cycle?
3. Can the solutions produced by ChatGPT serve students in carrying out validation – the final phase of the mathematical modelling cycle?

4 Method

As part of the course Mathematical Modelling, which is conducted in the seventh semester of the Applied Mathematics program at Faculty of Science, University of Kragujevac, students are expected, upon completion of the course, to be able to apply the principles of mathematical modelling and to formulate mathematical models in various fields of natural and social sciences. This course was offered for the first time in the 2023/2024 academic year. Only four students attended it during that academic year. After completing the theoretical and practical instruction, the students took the written part of the exam, in which they were expected to demonstrate practical knowledge of mathematical modelling. On that occasion, they solved tasks in which they were required to carry out the last four phases of mathematical modelling, specifically from mathematization to validation [1]. The test consisted of 6 tasks. The first task involved a simple theoretical consideration of an exponential growth model with a given constant. The second task focused on the application of a first-order differential equation to a bacterial population in a controlled environment, where students were required to construct an appropriate logistic differential equation using the given data. Using the model they created, they had to determine the size of the population at several specific time points from the initial moment, identify when the population would reach half the carrying capacity, and analyze what happens as the time variable tends to infinity. The third task was also a population model, but with constant relative rates of birth, death, and emigration over time. Students were asked to determine a particular solution to the problem for a given initial value, find the condition under which the model exhibits exponential growth, and identify the condition under which the population remains constant. In the fourth task, students were provided with tabular data related to time and population size and were required to construct a model that fits the given data. Based on the model, they were to explain the meaning of all parameters and their influence on the behavior of the model. The fifth problem dealt with mathematical modelling within decision theory – more specifically, game theory and the analysis of strategies in a competitive environment. The task required modelling the competitive relationship between two manufacturers, identifying the strategies of each player, and determining the optimal decision for the first company in response to various possible moves by its competitor. The final, sixth task focused on the application of linear programming, specifically optimizing production under resource constraints, with two objectives: minimizing production costs under the condition that a certain number of products be manufactured, and maximizing the number of products produced given limited resources.

We gave the same problems that the students had solved to ChatGPT to solve as well. Considering that we wanted to examine whether it could help students in terms of validating their answers by comparing them with the solution produced by ChatGPT, we used the free version of ChatGPT (GPT-4o). When evaluating the solutions provided by the students and the ChatGPT, we assessed the work according to the modelling cycle – that is, to what extent the problem solutions followed the given mathematical modelling approach. The study was conducted in July and August 2025.

5 Results

To address the research questions, we analyzed whether ChatGPT successfully solved the six tasks that the students had worked on. Specifically, for each task, we examined whether the following steps (Fig.1) were successfully carried out: translating the problem into a mathematical context, i.e., the process of mathematization; whether the problem-solving procedure was mathematically correct; interpreting the solution in accordance with the original context; whether and how the validation process was carried out.

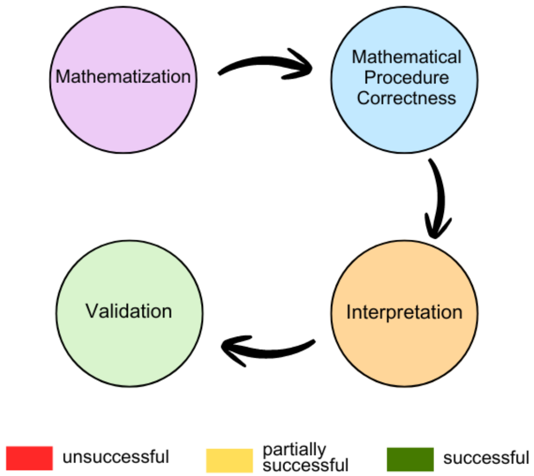


Fig. 1. Illustration of the evaluation of the mathematical modelling process

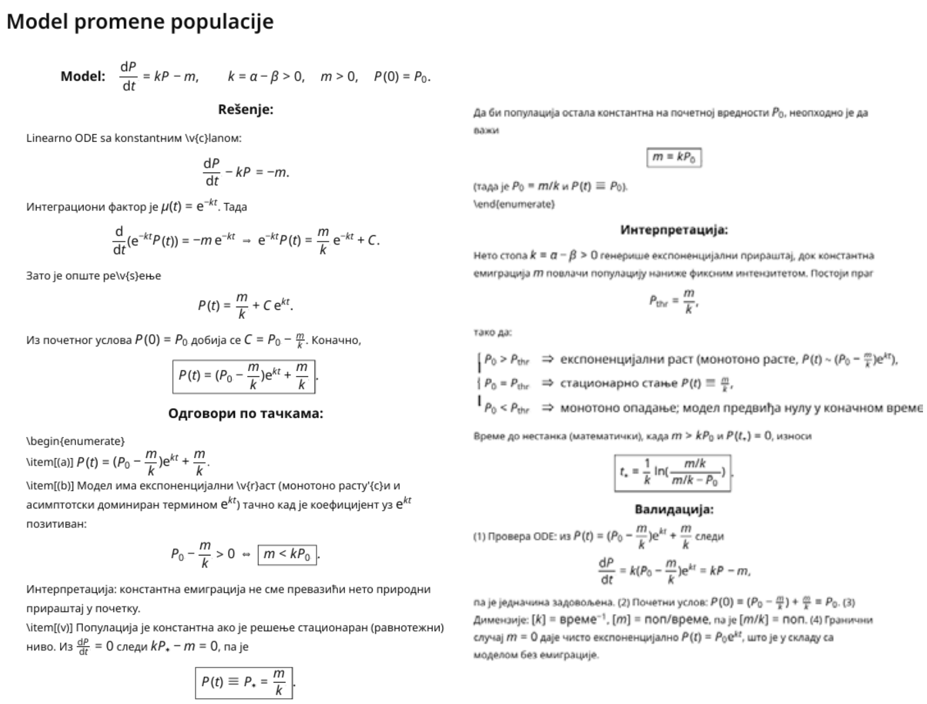
For each aspect and each task, one of three levels of success was assigned: successful, partially successful, and unsuccessful. Solutions categorized as unsuccessful were those in which serious conceptual or procedural errors were identified. The successful category included all solutions that were entirely correct, while partially successful referred to those parts of the solutions that were fundamentally correct but either lacked sufficient detail or precision, or contained a minor computational error in the mathematical result. The following table presents ChatGPT's performance in solving these tasks.

Table 1. ChatGPT's Performance in Solving Mathematical Modelling Tasks

Problem	Mathematization	Mathematical Result	Interpretation	Validation
1	successful	successful	successful	successful
2	successful	partially successful	partially successful	partially successful
3	successful	successful	successful	successful
4	partially successful	successful	partially successful	successful

5	successful	successful	partially successful	successful
6	partially successful	partially successful	partially successful	partially successful

At first glance, it can be seen that none of the four aspects in any of the six tasks was unsuccessful. A detailed analysis of the solutions produced by ChatGPT revealed that the first task was completed successfully. However, in the second task, which concerned a bacterial population model, it made an error in solving the model (although it should be emphasized that the chosen model was appropriate). Specifically, it proceeded with a numerical approach, which led to a computational error that subsequently affected the following steps of the solution. The third task was solved entirely successfully (Fig. 2), whereas in the fourth task – which, unlike the first three, was presented in a tabular form – ChatGPT constructed a model that was not the most suitable solution to the problem. In the fifth task, the processes of mathematization and solution implementation were adequate, but the interpretation of the solution within the appropriate context of game theory could have been better and more refined. The validation process, however, was carried out successfully. In the final task, which was from the field of linear programming, ChatGPT exhibited the greatest weaknesses, especially from the mathematical perspective, as it did not select the most appropriate model, and the objective function was also not well formulated. Nevertheless, the constraints were well defined, and the problem was solved correctly in accordance with them.



Model promene populacije

Model: $\frac{dP}{dt} = kP - m, \quad k = \alpha - \beta > 0, \quad m > 0, \quad P(0) = P_0.$

Rešenje:

Linearno ODE sa konstantним \v{c}lanом:

$$\frac{dP}{dt} - kP = -m.$$

Интеграциони фактор је $\mu(t) = e^{-kt}$. Тада

$$\frac{d}{dt}(e^{-kt}P(t)) = -m\,e^{-kt} \Rightarrow e^{-kt}P(t) = \frac{m}{k}e^{-kt} + C.$$

Зато је опште ре\v{s}ење

$$P(t) = \frac{m}{k} + C\,e^{kt}.$$

Из почетног услова $P(0) = P_0$ добија се $C = P_0 - \frac{m}{k}$. Коначно,

$$\boxed{P(t) = (P_0 - \frac{m}{k})e^{kt} + \frac{m}{k}}.$$

Одговори по тачкама:

\begin{enumerate}
\item[(a)] $P(t) = (P_0 - \frac{m}{k})e^{kt} + \frac{m}{k}$.
\item[(b)] Модел има експоненцијални \v{r}аст (монотоно расту'{c}и и асимптотски доминиран термином e^{kt}) тачно кад је коефицијент уз e^{kt} позитиван:

$$P_0 - \frac{m}{k} > 0 \Leftrightarrow \boxed{m < kP_0}.$$

Интерпретација: константна емиграција не сме превазићи нето природни прираштај у почетку.
\item[(v)] Популација је константна ако је решење стационаран (равнотежни) ниво. Из $\frac{dP}{dt} = 0$ следи $kP_* - m = 0$, па је

$$\boxed{P(t) \equiv P_* = \frac{m}{k}}.$$

Да би популација остала константна на почетној вредности P_0, неопходно је да важи

$$\boxed{m = kP_0}$$

(тада је $P_0 = m/k$ и $P(t) \equiv P_0$).
\end{enumerate}

Интерпретација:

Нето стопа $k = \alpha - \beta > 0$ генерише експоненцијални прираштај, док константна емиграција m повлачи популацију наниже фиксним интензитетом. Постоји праг

$$P_{thr} = \frac{m}{k},$$

тако да:

$P_0 > P_{thr} \Rightarrow$ експоненцијални раст (монотоно расте, $P(t) \sim (P_0 - \frac{m}{k})e^{kt}$),
$P_0 = P_{thr} \Rightarrow$ стационарно стање $P(t) \equiv \frac{m}{k}$,
$P_0 < P_{thr} \Rightarrow$ монотоно опадање; модел предвиђа нулу у коначном време

Време до нестанка (математички), када $m > kP_0$ и $P(t_*) = 0$, износи

$$\boxed{t_* = \frac{1}{k}\ln(\frac{m/k}{m/k - P_0})}.$$

Валидација:

(1) Провера ODE: из $P(t) = (P_0 - \frac{m}{k})e^{kt} + \frac{m}{k}$ следи

$$\frac{dP}{dt} = k(P_0 - \frac{m}{k})e^{kt} = kP - m,$$

па је једначина задовољена. (2) Почетни услов: $P(0) = (P_0 - \frac{m}{k}) + \frac{m}{k} = P_0$. (3) Димензије: $[k]$ = време$^{-1}$, $[m]$ = поп/време, па је $[m/k]$ = поп. (4) Гранични случај $m = 0$ даје чисто експоненцијално $P(t) = P_0 e^{kt}$, што је у складу са моделом без емиграције.

Fig. 2. The solution to the third task by ChatGPT in the context of mathematical modelling.

Regarding the students' work, all students solved the first and second tasks completely successfully. In the third task, three students fully carried out the appropriate procedures in all four steps, while one student completed the first two steps entirely successfully but only partially succeeded in interpreting and validating the solution. In the fourth task, two students successfully completed all steps. One student was only partially successful in the mathematization process and in interpreting the solution but fully succeeded in the other two steps. Another student had the same level of success as the previous one in the first three parts but entirely omitted the validation step. In the fifth task, two students completed the task entirely successfully, while the other two were successful in mathematization but only partially successful in the remaining three steps. In the final task, two students again solved it completely successfully. One was partially successful in all steps, while the fourth student was fully successful in the first three steps but partially successful in validation.

To more easily compare the performance of students and ChatGPT, we numerically evaluated their results. Each task was worth 4 points (1 point for each of the four parts of the solution). Each of the four aspects in every task was scored with 1 if that part was fully completed successfully, 0.5 for partially successful, and 0 for unsuccessful. The students scored 24, 22, 20.5, and 20 points, respectively. The arithmetic mean of the students' scores was 21.625 points, with a standard deviation of 1.80 points. Looking at ChatGPT's results on this test, it scored 19 points. Compared to the students' results, the Z-score corresponding to ChatGPT's score is -1.46. This score is relatively low but not extreme, as it falls between one and two standard deviations below the mean. This result indicates that the students still performed better at solving mathematical modelling tasks. On the other hand, some students could benefit from certain aspects of ChatGPT's solutions, particularly in interpreting and validating the results.

6 Discussion and Conclusion

By reviewing the students' work, it can be observed that their shortcomings in mathematical modelling are more pronounced in steps such as interpreting results within the initial context and validating the obtained solution, while weaknesses in mathematizing the problem and carrying out computations are less noticeable. These deficiencies—or factors influencing students' success in mathematical modelling—align with those identified by Sen Zeytun et al. [3], particularly referring to difficulties in mathematization, i.e., establishing connections between the real-world problem and the mathematical context, and vice versa.

When it comes to ChatGPT's performance, we see that it successfully carried out the mathematization process in 4 out of 6 tasks – meaning it understood the problem's context and correctly assigned mathematical meaning to the models. In the same number of cases, it was fully successful in solving mathematical problems, including constructing equations, solving differential equations, and addressing tasks from other mentioned fields. That ChatGPT is successful in solving first-order differential equations was also demonstrated by Koceska et al. [8]. In the remaining two cases, it was partially successful. On the other hand, Li et al. [11] pointed out the shortcomings of ChatGPT in solving linear programming problems compared to an LLM that had been additionally trained to handle that specific type of task. In our research as well, ChatGPT was partially successful when solving the linear programming task. Interestingly, it was only partially successful in both initial steps of the linear programming task. More concerning is that in 4 out of 6 cases, it was only partially successful in interpreting the results, while in the remaining two, it was fully successful. It should be noted that the tasks given to ChatGPT did not include explicit instructions regarding the mathematical modelling process, so future research should examine its solutions in more detail if more specific modelling-related instructions are provided. As for validation, ChatGPT was fully successful in 4 out of 6 cases, while in the remaining two, it was only partially successful.

Based on the obtained results, we can say that ChatGPT largely successfully solved mathematical modelling problems related to the application of differential equations, game theory, and linear programming – though these results could have been better. When comparing ChatGPT's results with those of the students, we observe that all four students achieved better results in terms of solution structure within the mathematical modelling cycle. However, it is also evident that certain parts of ChatGPT's solutions could still be useful to some students, particularly in verifying and validating problem solutions.

Obviously, this research has its limitations. The first limitation concerns the sample size, which stems from the fact that only four students were enrolled in the Mathematical Modelling course during the given academic year. Consequently, our findings cannot be generalized. Furthermore, in this research we examined the capabilities of only one tool - the free version of ChatGPT - because it is accessible to all students without additional subscription costs. These limitations also point to directions for future research. In addition to providing more detailed instructions when assigning tasks within the framework of the mathematical modelling cycle, future studies could also investigate the capabilities of other free tools, such as DeepSeekR1, as well as paid versions of widely adopted tools, including those offered by OpenAI, Claude Opus, Gemini, and others, and compare their performance on tasks completed by larger student samples. Furthermore, future research could focus on developing specialized LLM models designed to address and solve relevant problems in accordance with the principles of mathematical modelling, with the goal of providing higher-quality support to students.

Acknowledgments. This work was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia through the Agreement No. 451-03-65/2024-03/200122.

Disclosure of Interests. No potential conflict of interest was reported by the authors.

References

[1] Blum, W., & Leiss, D. (2007). How do students and teachers deal with modelling problems. *Mathematical modelling (ICTMA 12): Education, engineering and economics*, 222-231.
[2] Schukajlow, S., Kaiser, G., & Stillman, G. (2018). Empirical research on teaching and learning of mathematical modelling: A survey on the current state-of-the-art. *ZDM*, *50*(1), 5-18.
[3] Sen Zeytun, A., Cetinkaya, B., & Erbas, A. K. (2023). Why do prospective teachers have difficulties in mathematical modelling? Insights from their perspectives. *International Journal of Mathematical Education in Science and Technology*, *55*(1), 94–117. https://doi.org/10.1080/0020739X.2023.2171922
[4] Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., et al. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med. Educ. 9*, doi: 10.2196/45312
[5] Svičević, M., Milenković, A., Vučićević, N. & Stanković, M. (2025) Evaluating the Success of AI Tools in Supporting Student Performance in Mathematical Kangaroo Competition, *Computer Applications in Engineering Education 33*(4), https://doi.org/10.1002/cae.70063
[6] Milenković, A., Vučićević, N., & Svičević, M. (2025) Evaluating open ai tools o1 and o3-mini in solving high school problems from Serbian national mathematics competition, *The teaching of mathematics, XXVIII* (1), 14 - 29, DOI:10.57016/TM-EYRQ1676
[7] Helfrich-Schkarbanenko, A. (2023). Mathematik und ChatGPT: ein Rendezvous am Fuße der technologischen Singularität. Springer Spektrum.
[8] Koceska, N., Koceski, S., Lazarova, L. K., Miteva, M., & Zlatanovska, B. (2023). Can ChatGPT be used for solving ordinary differential equations. *Balkan Journal of Applied Mathematics and Informatics*, *6*(2), 103-114.
[9] Zhu, T., Li, X., & Zhang, W. (2023). Applying chatgpt-powered game theory in ransomware negotiations. *Authorea Preprints*.
[10] Huang, X., Shen, Q., Hu, Y., Gao, A., & Wang, B. (2024). Llms for mathematical modeling: Towards bridging the gap between natural and mathematical languages. *arXiv preprint arXiv:2405.13144*.
[11] Li, Q., Zhang, L., & Mak-Hau, V. (2023). Synthesizing mixed-integer linear programming models from natural language descriptions. *arXiv preprint arXiv:2311.15271*.