

## EXPLORING THE PERFORMANCE OF CHATGPT FOR NUMERICAL SOLUTION OF ORDINARY DIFFERENTIAL EQUATIONS

Saso Koceski , Natasa Koceska , Limonka Koceva Lazarova ,  
Marija Miteva\* , Biljana Zlatanovska 

Goce Delcev University, Stip (Macedonia)

*saso.koceski@ugd.edu.mk, natasa.koceska@ugd.edu.mk, limonka.lazarova@ugd.edu.mk,*

*\*Corresponding author: marija.miteva@ugd.edu.mk, biljana.zlatanovska@ugd.edu.mk*

*Received February 2024*

*Accepted September 2024*

### Abstract

This study aims to evaluate ChatGPT's capabilities in certain numerical analysis problem: solving ordinary differential equations. The methodology which is developed in order to conduct this research takes into account the following mathematical abilities (defined according to National Centre for Education Statistics): Conceptual Understanding, Procedural Knowledge, Problem Solving, and Application in Real-world Contexts. The outcomes demonstrate that ChatGPT's performed very well for the set tasks, and it also gives promising results for programming code generation, with certain limitations. The effectiveness and accuracy of the answers and solutions obtained by ChatGPT are related to the type of equation, i.e., how complex it is, and also with the instructions we give to ChatGPT. It also requires further improvement of the machine learning model and the ability to provide an explanation of how the output was obtained.

**Keywords** – Artificial intelligence (AI), Differential equations, Numerical solution, ChatGPT.

### To cite this article:

Koceski, S., Koceska, N., Lazarova, L.K., Miteva, M., & Zlatanovska, B. (2025). Exploring the performance of ChatGPT for numerical solution of ordinary differential equations. *Journal of Technology and Science Education*, 15(1), 18-34. <https://doi.org/10.3926/jotse.2709>

-----

## 1. Introduction

According to the definition of derivative of function, derivatives are based on certain change (change in the value of independent variable and the change in the value of function), thus because all the processes in the nature and also physical phenomena are based on particular change, wide range of them can be described with differential equations. This makes differential equations widely applicable in different fields (Abed & Khare, 2013; Lazarova, Stojkovikj, Miteva & Stojanova, 2021; Goyal, Kulczycki & Ram, 2022; Koceska & Koceski, 2022; Loginova, 2020; Mishi, Sabari, Amos, Egbogu, Kuje & Ojosipe, 2020; Momoniati, Myers, Banda & Charpin, 2012).

Although certain methods for analytical solution of different types of differential equations exist, also many numerical methods are used in solving differential equations. These methods are convenient when the differential equations are complex and cannot be solved analytically. Numerical methods are also used to simulate the behavior of physical systems and predict their performance under different conditions. The most often used methods in numerical analysis for determining solution of ordinary differential equations (ODEs) are (Atkinson, Han & Stewart, 2009): Picard's method, the Taylor series method, Euler's method, the Runge-Kutta method, Milne's method, the Adams-Bashforth method. Each of the mentioned methods has its own advantages and limitations. Which method is the most suitable for use, depends on the characteristics of the differential equation and the level of accuracy required for the problem at hand.

Finding numerical solution and writing numerical algorithms in order to solve numerically differential equations, is not always easy to do, and it requires an exceptional mathematical as well as programming skills (Rangelov, Dineva & Manolis, 2020). An attempt to overcome this problem is to use the current potential of Natural Language Processing (NLP) and Artificial Intelligence (AI), to automatically generate numerical solutions. For several years now, AI has been used to find solutions to various problems in different fields (Loshkovska & Koceski, 2015) such as: medicine (Kocev, Achkoski, Bogatinov, Koceski, Trajkovik, Stevanoski et al., 2018; Stojanov & Koceski, 2014; Trajkovik, Gjorgievska, Koceski & Kulev, 2014), biology (Stojanov, Mileva & Koceski, 2012; Terzievska, Todorov, Miteva, Doneva, Dyankova, Metodieva et al., 2020), engineering (Stamov, 2021), robotics (Koceska, Koceski, Beomonte, Trajkovik & Garcia, 2019; Koceski, Panov, Koceska, Beomonte & Durante, 2014; Koceski & Koceska, 2016), tourism (Koceski & Petrevska, 2012) as well as economy (Koceska & Koceski, 2014).

Among the various AI-based models, ChatGPT (ChatGPT website, n.d.) is nowadays one of the most successful and widely used. It is a language model developed by OpenAI (OpenAI website, n.d.), trained on a large corpus of text data (including mathematical problems and concepts), to follow an instruction and to provide detailed and accurate answers on various questions. A main advantage of ChatGPT over other existing language models, is its ability to learn and understand instructions given by the user and respond appropriately. ChatGPT actually behaves like a human who is constantly upgrading and learning new things. The ChatGPT performances have been explored on different tasks: for text generation (Khalil & Er, 2023; Fu, Teng, Georgaklis, White & Schmidt, 2023; Jiao, Wang, Huang, Wang, & Tu, 2023), text classification (Kuzman, Mozetic & Ljubescic, 2023; Amin, Cambria & Schuller, 2023; Huang, Kwak & An, 2023), code generation (Megahed, Chen, Ferris & Jones-Farmer, 2023; Sobania, Briesch, Hanna & Petke, 2023), quality assessment (Kocmi & Federmann, 2023; Wang, Liang, Meng, Shi, Li, Xu et al., 2023) etc.) and in various fields (medicine (Tu, Ma, & Zhang, 2023; Lederman, Lederman & Verspoor, 2022), healthcare (Nov, Singh & Mann, 2023), physics (West, 2023; Kortemeyer, 2023; Lehnert, 2023), education (Cotton, Cotton & Shipway, 2023; Tapalova & Zhiyenbayeva, 2022; Kumar & Boulanger, 2020; Choi, Hickman, Monahan & Schwarcz, 2023) etc.

In the field of mathematics, Frieder, Pinchetti, Griffiths, Salvatori, Lukasiewicz, Petersen et al. (2023) tested ChatGPT math capabilities on datasets that were publicly available and on hand-crafted ones (GHOSTS and miniGHOSTS - created using novel methodology). The results suggest that the mathematical knowledge and skills of an average math graduate student are higher than ChatGPT's math capabilities. Shakarian, Koyalamudi, Ngu and Mareedu (2023) explored the ChatGPT capabilities on Math Word Problems (MWP), and concluded that ChatGPT's capabilities varies significantly depending on the given task and on the requirement to demonstrate how the solution to that task was arrived at. Pardos and Bhandary (2023) used the Open Adaptive Tutoring system (OATutor) to examine the efficacy of generated answers by ChatGPT in learning algebra. According to the research 70% of answers passed manual verification and can contribute to positive learning outcomes. But these scores were lower than the scores of human prompts. In (Dao & Le, 2023) authors have been exploring ChatGPT's math capabilities answering questions with multiple-choice, used for the Vietnamese National High School Graduation Exam (VNHSGE). Questions were prepared for different subjects and different difficulty levels. The results show that ChatGPT performed best on SAT Math competition (with the 70% success rate),

followed by VNHSGE mathematics (58.8%). For the other exams the success rate, according to the authors, was lower. In his research, Borji (2023) investigate ChatGPT's failures grouped into several categories, including mathematics, logic, and reasoning, among others.

Although exploring ChatGPT capabilities and testing its application in various fields are the focus of many researchers nowadays, to the best of authors' knowledge, a prior study for exploring ChatGPT capabilities in finding numerical solution of differential equations has not been previously published. Koceska, Koceski, Lazarova, Miteva and Zlatanovska (2023) has conducted initial research to test the ChatGPT's performances in numerical analysis with an emphasis on determining numerical solution of ODEs. In this study, an evaluation methodology that will evaluate ChatGPT's performances in determining numerical solution of first and second order ODEs, from several aspects, is presented. It considers the free version of ChatGPT chat-bot program and the existing theories for learning and developing mathematical skills.

Namely, our methodology starts from the basic presumption that knowledge development is not a linear process; rather, it usually includes several phases among which there are: Conceptual Understanding, Procedural Knowledge, Problem Solving and Application in Real-world Contexts. Each of these phases contributes to development of specific abilities and skills. Proposed methodology tries to evaluate ChatGPT's developed knowledge about numerical solution of ordinary differential equations of first and second order, using carefully tailored questions based on clearly defined indicators.

In the rest of this paper the methodology and evaluation process are described, and the obtained results are presented and discussed.

## 2. Methodology

Development of mathematical knowledge refers to the process by which individuals acquire and advance their understanding of mathematical concepts, principles, and problem-solving techniques. Therefore, comprehending of any specific mathematical topic, usually includes several phases among which are:

- **Conceptual Understanding:** It involves grasping fundamental mathematical concepts and understanding their definitions.
- **Procedural Knowledge:** This aspect involves the ability to perform mathematical operations, algorithms, and procedures accurately and efficiently.
- **Problem Solving:** Mathematical knowledge is not merely about memorizing formulas or procedures but also about using that knowledge to evaluate a solution of different problems.
- **Application in Real-world Contexts:** The ultimate goal of mathematical learning is to apply acquired knowledge and skills to real-world problems. It usually involves critical thinking i.e., making conjectures, logical reasoning and justifying claims.

According to the existing theories for evaluation of developed mathematical knowledge in students, we have developed a methodology for evaluation of ChatGPT's knowledge about abilities to solve numerically ordinary differential equations of first and second order. Proposed methodology tries to challenge ChatGPT on the given topic, using carefully tailored prompts, trying to reveal its maturity in each of the previously mentioned phases.

### 2.1. Conceptual Understanding and Procedural Knowledge

In order to evaluate the conceptual understanding and procedural knowledge a test was built based on the indicators of the conceptual understanding and procedural knowledge ability given in Table 1.

The test for evaluation of conceptual understanding was composed of 25 multi-choice and essay type of questions while the test for evaluation of procedural knowledge contained 20 multi-choice questions (MCQ). Each MCQ consisted of a stem and options. If there was a need for an auxiliary information, it was usually included in the stem. The biggest constrain in the process of questions' definition was the fact

that they should be defined using text only (without images or any other graphical representations). Moreover, in order to avoid mismatches and problems with more complicated formulas we have used ChatGPT Equation Renderer extension for Chrome that enables correct display of equations in ChatGPT using Latex notation. However, one of the most challenging tasks was formulation of multi-choice options that include one correct answer and multiple distractors. For definition of effective distractors, we have applied the domain knowledge and have applied common methodology that focuses on a selection of misconceptions or errors in thinking, reasoning, and problem solving by evaluating student answers on these or similar problems in the past.

	<b>Conceptual understanding indicators</b>	<b>Procedural understanding indicators</b>
1.	Ability to restate the concept of numerical methods for solving ordinary differential equations	Ability to use and utilize certain numerical methods
2.	Ability to recognize, distinguish and classify the methods and approaches	Ability to apply numerical methods in particular situations

Table 1. Conceptual and procedural test indicators

The answers generated by ChatGPT were evaluated independently by three experts on a 5 points Lickert scale, where 5 is the highest grade. Final mark for every problem was obtained as an average of their grades. So, not only correctly selected option but, also the rational explanation behind it was evaluated.

## 2.2. Problem Solving Capacities

The ability to solve problems enables individuals to analyse complex situations, devise strategies, and arrive at logical solutions. In mathematics, problem-solving goes beyond mere calculation; it involves critical thinking, creativity, and perseverance. For evaluation of ChatGPT skills for numerical solutions of ODE a specific methodology for construction of problems dataset was developed and applied. The steps of the research process are shown on Figure 1.

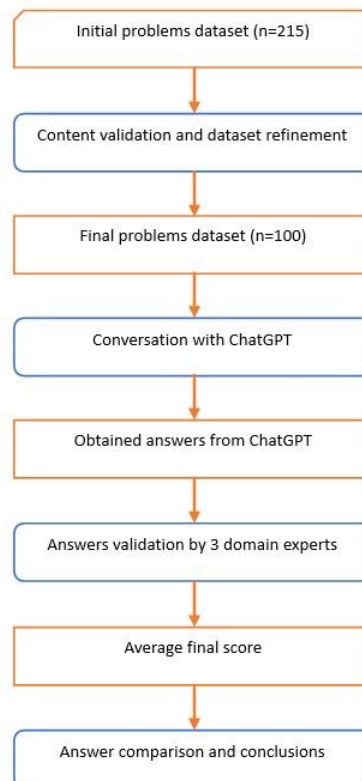


Figure 1. Flowchart of the evaluation methodology

For the purpose of this research, we have extracted ordinary differential equations of first and second order from various academic books and textbooks (Lebl, 2022; Strang, 2022; Nagy, 2021; Nagle, Edward & Snider, 2012; Tenenbaum & Pollard, 1963; Bayen, Kong & Siau, 2020; Johansson, 2015), and also written our own equations, in order to create the initial dataset. The three experts have reviewed this dataset and reduced the content to 100 equations (50 of them were first order and other 50 second order equations).

After creation of the initial dataset, we have started the conversation with ChatGPT. The answers generated by ChatGPT were evaluated by our experts, independently, on a scale from 1 to 5, where 5 is the highest grade. Final mark for every problem was obtained as an average of their grades.

### 2.3. Capability to Apply the Knowledge in Real-World Contexts

Contemporary theories of learning, state that the process of learning goes far beyond simple acquiring of information. It aims at development and nurturing of thinking abilities at higher level. This means the capacity to think more than remembering facts and emphasizing meaning to obtain solutions to problems by analysing, evaluating, and creating. According to Polya (2004) problem-solving is the highest level of higher-order thinking ability by combining creative thinking with critical thinking. Capacity to apply the gained knowledge for solution of variety of mathematical problems is usually considered one of the main indicators and factors that reflects the quality of learning. However, solving real-world problems, goes one step beyond as it requires establishment of strong mathematical connections and using appropriate math concept or apparatus that fits properly with conditions given in the problem. Therefore, in order to evaluate the capabilities of ChatGPT on solving real-life problems using numerical methods for solution of ODE, we chose the problem described by the Newton's law of cooling.

Newton's law of cooling is a physics law stating that when an object with temperature  $T(t)$  at time  $t$  is placed in a surrounding with temperature  $T_s$ , then the rate of change of  $T$  at time  $t$  is proportional to  $T(t) - T_s$ . Therefore, this process could be described with the following differential equation:

$$dT/dt = -k(T - T_s) \quad (1)$$

Constant  $k$  is known as decay constant. In this context,  $k > 0$ , since the temperature of the object must decrease if  $T > T_s$ , or increase if  $T < T_s$ .

To solve real-world problems described with this law, it is necessary to establish strong mathematical connections and to pass through series of processes such as: understanding the problem, modelling the problem, proposing methodologies or solution(s), application, and drawing conclusions.

The following methodology was applied: 1) clear presentation of the problem, 2) formulating the hypothesis, 3) proposing solution(s), 4) testing the solution, 5) drawing conclusions.

Simple conversation with ChatGPT was performed. Every answer was evaluated independently by three human experts using the following Likert (1-5; 1 being lowest and 5 highest value) scoring guidelines:

Score	Procedural knowledge indicators
1	No answer
2	Limited understanding of the problem, inability to recognize the components, connections are not clear, incomplete solutions
3	Fair understanding of the problem, able to recognize the components, but with less understanding, connections are not clear, incomplete solutions
4	Complete understanding of the problem, able to recognize the components with clear understanding, connections are not complete, incomplete solutions
5	Complete understanding of the problem, able to recognize the components with complete understanding, connections are fully established, solutions are complete and systematic

Table 2. List of scores and corresponding criteria for evaluation of real-world problem-solving capabilities

### 3. Evaluation

Entire evaluation process was conducted using the free research preview version of ChatGPT (release July 20, 2023). The evaluation process was conducted manually through ChatGPT's user interface using Chrome browser.

All the questions aimed at evaluation of ChatGPT's conceptual understanding and procedural knowledge were posed to ChatGPT in a single session. The stem and the options for each question were submitted as a prompt to ChatGPT. Answers on multi-choice questions, obtained from ChatGPT, besides the chosen option contained a short rational explanation. While the essay questions were answered in a narrative format. Part of the conversation can be seen in the following figures:

**S** The ordinary differential equation  $dy/dx = -3x + 2$ , with  $x(0) = 1$  is to be solved using the forward Euler method. What is the largest time step that can be used to solve the equation without making the numerical solution unstable?

**AI** To ensure numerical stability while using the forward Euler method, the time step ( $h$ ) must satisfy the stability condition. For the forward Euler method, the stability condition requires that  $h$  must be less than a certain value for the numerical solution to remain stable.

The stability condition for the forward Euler method is given by:

$$h < \frac{2}{\lambda_{\max}}$$

where  $\lambda_{\max}$  is the largest eigenvalue of the coefficient matrix of the differential equation. In this case, the given differential equation is  $\frac{dy}{dx} = -3x + 2$ .

The coefficient matrix is simply the coefficient of  $y$  in the equation, which is  $-3$  in this case. So the largest eigenvalue is  $-3$ .

Now, we can calculate the largest time step ( $h_{\max}$ ) that can be used to ensure numerical stability:

$$h_{\max} = \frac{2}{\lambda_{\max}} = \frac{2}{-3} = -\frac{2}{3}$$

However, time step values should always be positive, so the largest positive time step that can be used for numerical stability is  $h_{\max} = \frac{2}{3}$ .

Therefore, the largest time step that can be used to solve the given differential equation using the forward Euler method without making the numerical solution unstable is  $h = \frac{2}{3}$ .

Figure 2. Part of the conversation for evaluation of procedural knowledge

**S** In the equation for Runge-Kutta method  $y(n+1) = y(n) + hT4(x(n), y(n), h)$ , the expression for  $T4$  is:

- (1)  $1/6(2k1 + 2k2 + 2k3 + 2k4)$
- (2)  $1/6(k1 + 2k2 + 2k3 + k4)$
- (3)  $1/6(2k1 + k2 + 2k3 + k4)$
- (4)  $1/6(k1 + k2 + k3 + k4)$

**AI** The expression for  $T4$  in the Runge-Kutta method is:

$$(2) \frac{1}{6} (k1 + 2k2 + 2k3 + k4)$$

In the fourth-order Runge-Kutta method, to obtain the next approximation  $y(n + 1)$  from the current approximation  $y(n)$ , the following intermediate values are calculated:

$$k1 = h \cdot f(x(n), y(n))$$

$$k2 = h \cdot f\left(x(n) + \frac{h}{2}, y(n) + \frac{k1}{2}\right)$$

Figure 3. Sample conversation for evaluation of conceptual knowledge – multi-choice question



For evaluation of problem-solving capabilities, each of the selected problems was subject of separate conversation. Conversations with ChatGPT consists of the following parts:

- A prompt containing definition of the problem i.e., the differential equation, initial conditions, identification of particular numerical method (Picard's method, Taylor's series method, Euler's method, Milne's method, Runge-Kutta method or Adams-Bashforth method) that should be used, as well as requirement to find the value of the function at given point;
- A prompt with a requirement for generation of programming code corresponding to the solution.

All the conversations with ChatGPT were conducted within the same session (one after another). Regardless of the method used, ChatGPT was asked to generate a Python code. Once the code was generated, it was transferred by the user to Google Colaboratory in order to be evaluated. Part of the conversation can be seen in the following figures:

**S** Can you please find me the numerical solution of the ODE  $y'=x+y$  using Taylor's series of fourth order. Find the values of  $y$  at  $x=0.1$  and  $x=0.2$ , given that  $y(0)=1$ . Please elaborate step by step

**AI** Sure! Let's find the numerical solution of the ODE  $y' = x + y$  using Taylor's series of fourth order and calculate the values of  $y$  at  $x = 0.1$  and  $x = 0.2$ , given that  $y(0) = 1$ .

Given ODE:  $y' = x + y$

Step-by-step procedure for Taylor's series expansion of fourth order:

1. Initial condition:  $y(0) = 1$
2. Taylor Series Expansion:  
We'll expand  $y$  in a Taylor series around  $x = 0$ :  
$$y(x) = y(0) + y'(0) * x + (y''(0) / 2!) * x^2 + (y'''(0) / 3!) * x^3 + (y^{(4)}(0) / 4!) * x^4 + \dots$$
3. Derivatives of  $y$ :  
$$y' = x + y$$
  
$$y'' = 1 + y'$$
  
$$y''' = y'' \text{ (since } y''' \text{ is equal to the second derivative } y'' \text{ in this case)}$$
  
$$y^{(4)} = y''' \text{ (since } y^{(4)} \text{ is equal to the third derivative } y''' \text{ in this case)}$$
4. Substitute initial condition:  
$$y(0) = 1$$
  
$$y'(0) = 0 \text{ (since } y'(0) = x + y = 0 + 1 = 1)$$
  
$$y''(0) = 1 + y'(0) = 1 + 1 = 2$$
5. Approximations for  $y$  at subsequent points:  
Using the Taylor series expansion up to the fourth order, we can approximate  $y$  at  $x = h$ , where  $h$  is the step size.  
$$y(h) \approx y(0) + y'(0) * h + (y''(0) / 2!) * h^2 + (y'''(0) / 3!) * h^3 + (y^{(4)}(0) / 4!) * h^4$$
  
$$y(h) \approx 1 + 0 * h + (2 / 2) * h^2 + (2 / 6) * h^3 + (2 / 24) * h^4$$
  
Similarly, we can obtain approximations for  $y$  at other points:  $y(2h)$ ,  $y(3h)$ ,  $y(4h)$ , and so on.
6. The general formula for the  $n$ th approximation:  
$$y(nh) \approx 1 + 0 * nh + (2 / 2) * (nh)^2 + (2 / 6) * (nh)^3 + (2 / 24) * (nh)^4$$
7. Numerical solution and values of  $y$  at  $x = 0.1$  and  $x = 0.2$ :  
Choose an appropriate step size  $h$  (e.g.,  $h = 0.1$ ).  
Calculate  $y$  at  $x = 0.1$  and  $x = 0.2$  using the formula obtained in step 6.

Let's proceed with the calculations:

1. For  $x = 0.1$ :  
$$y(0.1) \approx 1 + 0 * 0.1 + (2 / 2) * (0.1)^2 + (2 / 6) * (0.1)^3 + (2 / 24) * (0.1)^4$$
2. For  $x = 0.2$ :  
$$y(0.2) \approx 1 + 0 * 0.2 + (2 / 2) * (0.2)^2 + (2 / 6) * (0.2)^3 + (2 / 24) * (0.2)^4$$

Now, let's perform the calculations:

1. For  $x = 0.1$ :  
$$y(0.1) \approx 1 + 0.01 + 0.00333333 + 0.000166667 \approx 1.1145$$
2. For  $x = 0.2$ :  
$$y(0.2) \approx 1 + 0.04 + 0.01333333 + 0.000666667 \approx 1.248$$

So, the approximate values of  $y$  at  $x = 0.1$  and  $x = 0.2$  are 1.1145 and 1.248, respectively. These are the numerical solutions obtained using Taylor's series expansion of fourth order with the given initial condition  $y(0) = 1$ .

Figure 4. Sample conversation demonstrating the problem-solving capabilities – part 1

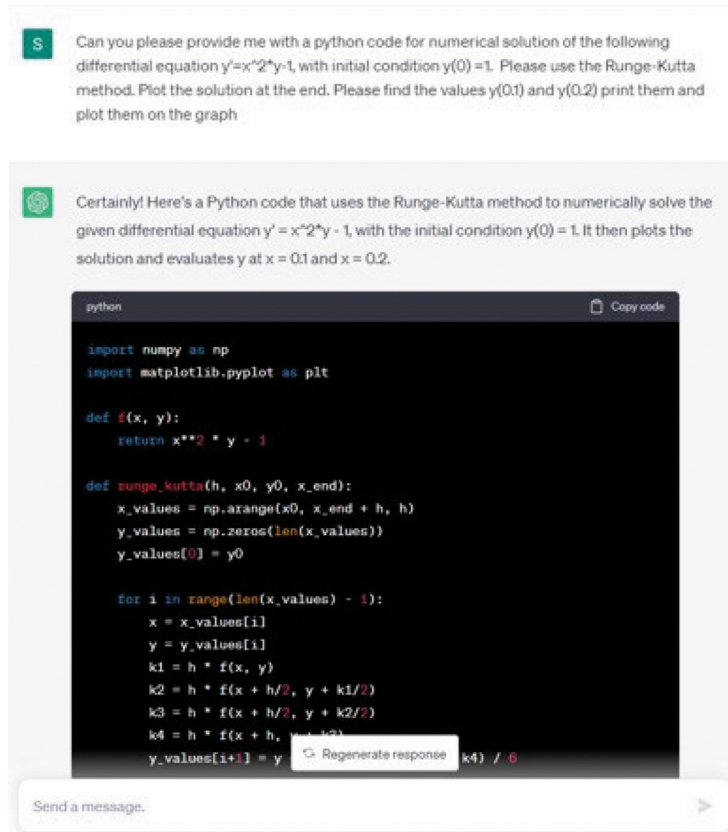


Figure 5. Sample conversation demonstrating the problem-solving capabilities – part 2

In order to evaluate the problem-solving skills on real-world problems, we have conducted the conversation with ChatGPT in a separate session, following the defined methodology. We have prepared multiple prompts that were posed to ChatGPT in a sequential order. During the conversation we gave clear feedback and incentives for good answers. Part of this conversation is presented on the figures in the following:

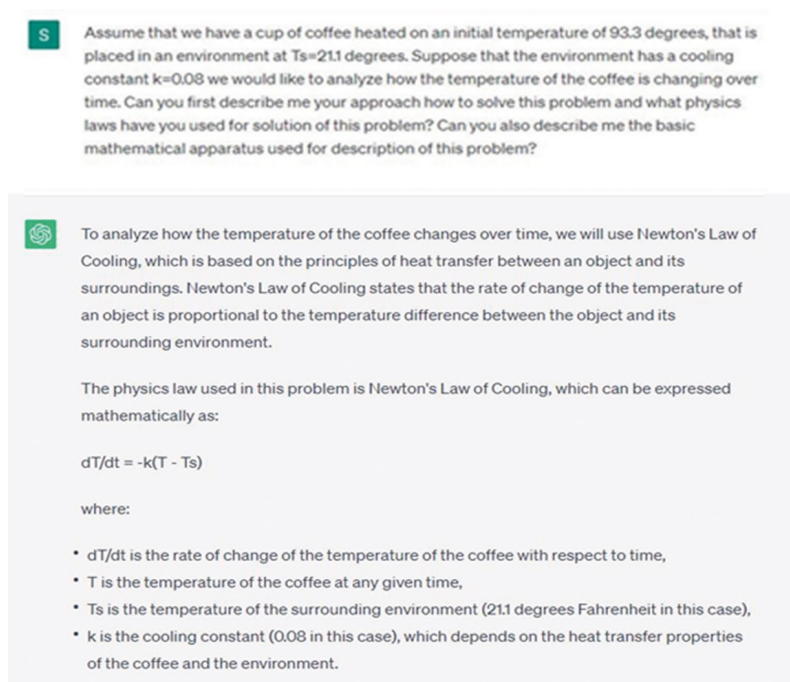


Figure 6. Beginning of the conversation that should indicate basic conceptual knowledge, procedural knowledge and connection of concepts – part 1



Mathematically, this is a first-order ordinary differential equation (ODE) with the initial condition:

$$T(0) = 93.3$$

This ODE can be solved analytically using integration, as shown in the previous response. The analytical solution is:

$$T(t) = T_s + (T_0 - T_s) * e^{(-kt)}$$

where:

- $T(t)$  is the temperature of the coffee at time  $t$ ,
- $T_0$  is the initial temperature of the coffee (93.3 degrees Fahrenheit),
- $T_s$  is the surrounding temperature (21.1 degrees Fahrenheit),
- $k$  is the cooling constant (0.08).

The analytical solution provides an exact expression for the temperature of the coffee as a function of time. It gives us a clear understanding of how the temperature changes over time and how it approaches the surrounding temperature.

Apart from the analytical solution, numerical methods like the fourth-order Runge-Kutta method and Euler's method can be used to approximate the solution of this ODE. These numerical methods involve dividing the time interval into small steps and iteratively calculating the temperature at each step. By using smaller time steps, we can achieve more accurate numerical approximations.

The basic mathematical apparatus used for the description of this problem involves solving the first-order ODE using integration techniques to obtain an analytical solution. Regenerate

Figure 7. Beginning of the conversation that should indicate basic conceptual knowledge, procedural knowledge and connection of concepts – part 2

**S** Having in mind the problem with the cooling of the coffee we gave you before, can you please solve me the proposed ODE analytically? Don't forget that the initial temperature was 93.3. Please find all the coefficients and write down the final analytical solution i.e. the formula with calculated coefficients?

**AI** Certainly! Let's solve the ordinary differential equation (ODE) for the cooling of the coffee analytically.

**Given Information:**

- Initial temperature ( $T_0$ ) = 93.3 degrees Fahrenheit
- Surrounding temperature ( $T_s$ ) = 21.1 degrees Fahrenheit
- Cooling constant ( $k$ ) = 0.08

**ODE:**  
The ODE describing the cooling of the coffee is given by Newton's Law of Cooling:

$$dT/dt = -k(T - T_s)$$

Figure 8. Sample conversation that should test the capabilities to apply the gained knowledge on a corresponding mathematical problem i.e., solution of first-order ODE – part 1

**Analytical Solution:**  
 To solve the ODE analytically, we can separate variables and integrate both sides:

$$\int dT / (T - T_s) = -k \int dt$$

Integrating the left side gives:

$$\ln |T - T_s| = -kt + C$$

where C is the integration constant.

Now, let's solve for T:

$$|T - T_s| = e^{-kt + C}$$

Since  $T(0) = 93.3$ , we can substitute  $t = 0$  and  $T = 93.3$ :

$$|93.3 - 21.1| = e^{-(0) + C}$$

$$72.2 = e^C$$

$$C = \ln(72.2)$$

Therefore, the analytical solution for  $T(t)$  is:

$$T(t) = T_s + (T_0 - T_s) * e^{-kt}$$

where:

- $T_s = 21.1$  degrees Fahrenheit (surrounding temperature)
- $T_0 = 93.3$  degrees Fahrenheit (initial temperature)
- $k = 0.08$  (cooling constant)

Substituting the values:

$$T(t) = 21.1 + (93.3 - 21.1) * e^{-0.08 * t}$$

So, the final analytical solution for the temperature of the coffee at any time  $t$  is:

$$T(t) = 21.1 + 72.2 * e^{-0.08 * t}$$

Figure 9. Sample conversation that should test the capabilities to apply the gained knowledge on a corresponding mathematical problem i.e., solution of first-order ODE – part 2

**S** You've mentioned that the ODE describing this problem could also be solved numerically using for e.g. Euler's method and Runge-Kutta method of order 4. Can you please provide me Python code for numerical solution of this problem using both methods and a time step of  $h=0.5$ . Please calculate the temperature with precision of 10 decimal places at every step from time unit 0 to 10. Please also include the analytical formula in the code and calculate the "exact temperature" in each of the steps. Please plot the solution curve and mark the found points by the analytical solution and using both numerical methods with different circles or dots on the same plot. Please elaborate how did you find the solution?

**S** Sure! To solve the cooling problem numerically using Euler's method and the fourth-order Runge-Kutta method, we can follow these steps:

1. Define the given parameters: initial temperature ( $T_0$ ), surrounding temperature ( $T_s$ ), cooling constant ( $k$ ), time step ( $h$ ), and the time range (from 0 to 10).
2. Implement the analytical solution formula to calculate the exact temperature at each time step.
3. Implement Euler's method and the fourth-order Runge-Kutta method to numerically approximate the temperature at each time step.
4. Calculate the temperature using both numerical methods and the analytical solution for each time step.
5. Plot the solution curves and mark the calculated points using the analytical solution, Euler's method, and the fourth-order Runge-Kutta method on the same plot.

Here's the Python code to achieve this:

```
python
import numpy as np
import matplotlib.pyplot as plt

# Given parameters
T0 = 93.3
Ts = 21.1
k = 0.08
h = 0.5
```

Figure 10. Further conversation that aimed at proving that the connections among concepts are fully established and provided solution is complete

```

ChatGPT.ipynb
File Edit View Insert Runtime Tools Help Last saved at 2:21AM

+ Code + Text
10 = 93.3 # initial temperature (degrees Fahrenheit)
k = 0.08 # Cooling constant

# Time step
h = 0.5

# Time array from 0 to 10 with step h
time_array = np.arange(0, 10.5, h)

# Initialize arrays to store results
euler_temperature = np.zeros(len(time_array))
rk4_temperature = np.zeros(len(time_array))
exact_temperature = np.zeros(len(time_array))

# Analytical solution function
def analytical_solution(t):
    return Ts + (T0 - Ts) * np.exp(-k * t)

# Numerical solution using Euler's method
def euler_method():
    euler_temperature[0] = T0
    for i in range(1, len(time_array)):
        euler_temperature[i] = euler_temperature[i - 1] - k * (euler_temperature[i - 1] - Ts) * h

# Numerical solution using 4th order Runge-Kutta method
def runge_kutta_method():
    rk4_temperature[0] = T0
    for i in range(1, len(time_array)):
        k1 = -k * (rk4_temperature[i - 1] - Ts)
        k2 = -k * (rk4_temperature[i - 1] + 0.5 * h * k1 - Ts)
        k3 = -k * (rk4_temperature[i - 1] + 0.5 * h * k2 - Ts)
        k4 = -k * (rk4_temperature[i - 1] + h * k3 - Ts)

        rk4_temperature[i] = rk4_temperature[i - 1] + (h / 6) * (k1 + 2 * k2 + 2 * k3 + k4)

# Calculate exact temperature for each time step using the analytical solution
for i, t in enumerate(time_array):
    exact_temperature[i] = analytical_solution(t)

# Calculate numerical solutions using Euler's method and Runge-Kutta method
euler_method()
runge_kutta_method()

# Plot the solution curve
plt.figure(figsize=(10, 6))
plt.plot(time_array, exact_temperature, label='Exact Temperature', color='blue', linewidth=2)
plt.plot(time_array, euler_temperature, label='Euler's Method', marker='o', markersize=5, linestyle='dashed', color='green')
plt.plot(time_array, rk4_temperature, label='Runge-Kutta Method', marker='s', markersize=5, linestyle='dotted', color='red')

plt.xlabel('Time (seconds)')
plt.ylabel('Temperature (Fahrenheit)')
plt.title('Newton's Law of Cooling - Numerical and Analytical Solutions')
plt.legend()
plt.grid(True)

```

Figure 11. Generated programming code was executed in Google Collaboratory without any human refinement

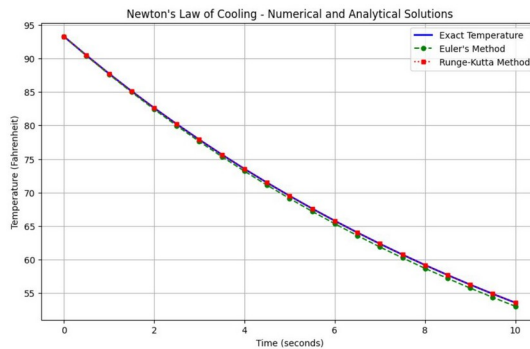


Figure 12. Graphical representation after code execution in Google Collaboratory clearly indicates the analytical solution as well as the results obtained with both numerical methods

Excellent. When we executed your code we obtained the following results

Time (s)	Exact Temp (F)	Euler Temp (F)	Runge-Kutta Temp (F)
0.0	93.3000000000	93.3000000000	93.3000000000
0.5	90.4689975068	90.4120000000	90.4689975680
1.0	87.7490002091	87.6395200000	87.7490003267
1.5	85.1356555310	84.9779392000	85.1356557005
2.0	82.6247815634	82.4228216320	82.6247817805
2.5	80.2123603722	79.9699087667	80.2123606330
3.0	77.8945315690	77.6151124161	77.8945318697
3.5	75.6675861331	75.3545079194	75.6675864701
4.0	73.5279604767	73.1843276026	73.5279608468
4.5	71.4722307423	71.1009544985	71.4722311423
5.0	69.4971073238	69.1009163186	69.4971077508
5.5	67.5994296022	67.1808796658	67.5994300535
6.0	65.7761608884	65.3376444792	65.7761613614
6.5	64.0243835634	63.5681387000	64.0243840558
7.0	62.3412944099	61.8694131520	62.3412949193
7.5	60.7242001260	60.2386366260	60.7242006504
8.0	59.1705130159	58.6730911609	59.1705135533
8.5	57.6777468488	57.1701675145	57.6777473974
9.0	56.2435128803	55.7273608139	56.2435134384
9.5	54.8655160301	54.3422663813	54.8655165961
10.0	53.5415512093	53.0125757261	53.5415517817

Based on these results can you please tell me which method gives better approximation?

Figure 13. Another challenge to test the deep reasoning over the obtained results – part 1

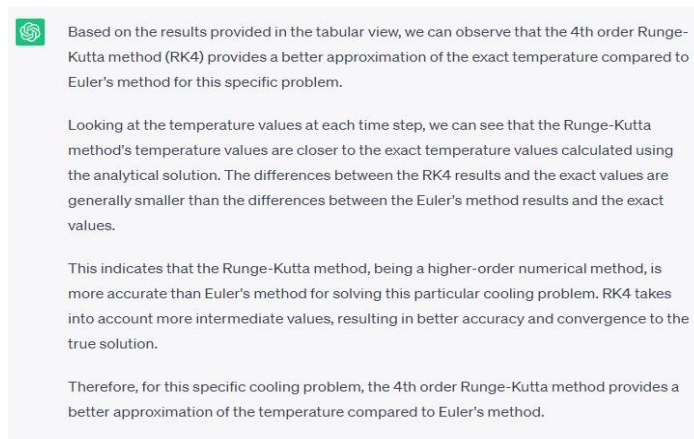


Figure 14. Another challenge to test the deep reasoning over the obtained results – part 2

#### 4. Results and Discussion

Answers on the questions aimed at evaluation of ChatGPT's conceptual knowledge demonstrated strong ability to identify all the facts related to the concepts of numerical methods for solution of ODE, restate the known principles and methods, clearly distinguish correct items and classify items based on the known concepts. All selected answers on the multi-choice questions were plausibly explained and clear connections between the concepts were given. This was somehow expected considering its predilection for natural language processing.

This conclusion is correlated with the overall average mark given by the domain experts of 4.8 on the scale from 1-5, 1 being smallest and 5 being highest mark. The common opinion is that ChatGPT knew all the arguments contained in the questions. Answers on the questions aimed at testing procedural knowledge of ChatGPT together with the given explanation of the choices made, were assessed by the domain experts with an overall average mark of 4.2, on the scale from 1-5, 1 being smallest and 5 being highest mark. Qualitative analysis of the answers revealed that in most of the cases ChatGPT was aware of the methods and algorithms, as well as how to apply them in a correct way. This was true even for more difficult questions where multiple connections were necessary.

For evaluation of ChatGPT's problem solving skills and ability to solve numerically ODE of first and second order the analysis was focused on multiple numeric methods. Figure 15 and Figure 16 depict ChatGPT's performance for the conducted research. The figures are box plots, where the x-axis reflects the methods used for finding numerical solution, while the y-axis reflects the grades given by domain experts. The plots show that the Runge-Kutta method gives the best results, in both cases (for first and second order ODEs). The center of the distributions is the highest for this plot, while the variability is lowest.

This means that the results of the experts' evaluation show the higher consistency about ChatGPT responses for the solution obtained by the Runge-Kutta method. They all agree that the answers given by Chat GPT using this method, is the most accurate and reliable.

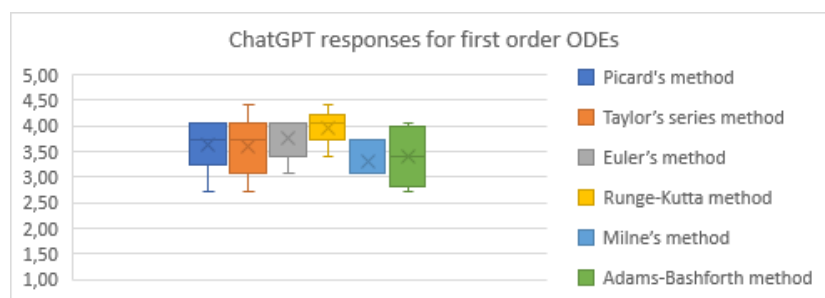


Figure 15. Grades from the experts regarding ChatGPT responses to solutions of first order ODEs



Figure 16. Grades from the experts regarding ChatGPT responses to solutions of second order ODEs

Despite the relatively satisfying results for problem solving capabilities, ChatGPT sometimes showed unsurprisingly inconsistent behavior when it comes to basic arithmetic operations, as well as in procedures involving simple items rearrangement. Moreover, we are not excluding the fact that the results may depend on the difficulty of the selected problems as well as on prompt engineering process. In this process we were mainly following the OpenAI instructions for step-by-step guidance.

The overall performance of ChatGPT for solving real-life problems (in our case connected to Newton's Law of cooling) was rated by the experts with an average mark of 4.1. Answers given by ChatGPT on various questions demonstrated that it completely understands the problem. Obtained answers are proving its ability to identify the elements with complete understanding and to establish full connections between mathematical concepts connected to numerical solution of ODE. In most of the cases obtained solutions were complete and systematic. However, the evaluation revealed one strange fact. Namely, ChatGPT failed to perform correctly some of the basic operations with floating point decimals when it was asked to do it with higher precision. So, after correct analytical solution of the first order ODE describing the given real-world problem, it was asked to calculate the "exact values" for the temperature in the moments of time  $t$  from 0 to 10 with a step of 0.5. Analyzing the obtained answers, we discovered that simple math operations like multiplication, division and exponential functions applied over floating-point numbers with higher precision including the Euler's constant, are producing wrong results.

According to the results, we can state that Chat GPT can deliver a comprehensive response in solving ODEs, accompanied by explicit mathematical and logical justifications.

## 5. Conclusions

The process of development of mathematical knowledge is usually considered one of the fundamental aspects of human education and intellectual growth. Each mathematic topic is not merely a subject to be memorized but a dynamic field that nurtures critical thinking, problem-solving skills, and logical reasoning.

Having in mind the developed theories for learning and acquiring and enlarging mathematical knowledge on different areas and considering the latest achievements in the field of AI, especially the public availability of ChatGPT, we have explored its capabilities in evaluating numerical solution of ordinary differential equations taking into account several aspects such as: Conceptual Understanding, Procedural Knowledge, Problem Solving and Application in Real-world Contexts.

In this research, ChatGPT's capabilities for numerically solving ordinary differential equations (ODEs) of first and second order were assessed using custom methodology. Additionally, its performance was evaluated in tackling a real-life problem based on Newton's law of cooling. Although the process of evaluation was limited with the fact that the used free and publicly available version of ChatGPT cannot be provided with graphical inputs and the results also depend on the process of prompt engineering, obtained results suggest that ChatGPT gained significant conceptual and procedural knowledge for solving numerically ODEs as well as generating programming code.



Moreover, it has demonstrated clear problem-solving skills capabilities. However, despite the fact that it copes very well with more complex problems, there are some inconsistencies when it comes to application of basic arithmetic operations and procedure.

Because the development of AI technology is rapidly increasing nowadays, as well as its use, it is expected that the application of AI for solution of specific mathematical problems, that require deep understanding and knowledge, will become more prevalent and effective in the future.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### References

- Abed, A.I., & Khare, R.K. (2013). Differential equations and their applications. *International Journal of Scientific Engineering and Technology Research*, 2(19), 2208-2217.
- Amin, M.M., Cambria, E., & Schuller, B.W. (2023). *Will affective computing emerge from foundation models and general ai? A first evaluation on ChatGPT*. arXiv preprint arXiv:2303.03186.
- Atkinson, K., Han, W., & Stewart, D. (2009). *Numerical Solution of Ordinary Differential Equations*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118164495>
- Bayen, A., Kong, O., & Siau, T. (2020) *Python Programming and Numerical Methods: A Guide for Engineers and Scientists*. Elsevier Science.
- Borji, A. (2023). *A categorical archive of ChatGPT failures*. arXiv preprint arXiv:2302.03494. <https://doi.org/10.21203/rs.3.rs-2895792/v1>
- ChatGPT website (n.d.). Available at: <https://chat.openai.com/>
- Choi, J.H., Hickman, K.E., Monahan, A., & Schwarcz, D. (2023). ChatGPT goes to law school. *71 Journal of Legal Education*, 387. <https://doi.org/10.2139/ssrn.4335905>
- Cotton, D.R., Cotton, P.A., & Shipway, J.R. (2023). Chatting and Cheating. Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. <https://doi.org/10.35542/osf.io/mrz8h>
- Dao, X.Q., & Le, N.B. (2023). *Investigating the Effectiveness of ChatGPT in Mathematical Reasoning and Problem Solving: Evidence from the Vietnamese National High School Graduation Examination*. arXiv preprint arXiv:2306.06331.
- Frieder, S., Pinchetti, L., Griffiths, R.R., Salvatori, T., Lukasiewicz, T., Petersen, P.C. et al. (2023). *Mathematical capabilities of ChatGPT*. arXiv preprint arXiv:2301.13867.
- Fu, Q., Teng, Z., Georgaklis, M., White, J., & Schmidt, D.C. (2023). *NL2CMD: An Updated Workflow for Natural Language to Bash Commands Translation*. arXiv preprint arXiv:2302.07845.
- Goyal, N., Kulczycki, P., & Ram, M. (2022). *Differential Equations in Engineering: Research and Applications*. CRC/Taylor & Francis. <https://doi.org/10.1201/9781003105145>
- Huang, F., Kwak, H., & An, J. (2023). Is ChatGPT better than human annotators? Potential and limitations of ChatGPT in explaining implicit hate speech. In *Companion Proceedings of the ACM Web Conference 2023*. Association for Computing Machinery (294-297). New York. <https://doi.org/10.1145/3543873.3587368>



- Jiao, W., Wang, W., Huang, J.T., Wang, X., & Tu, Z.P. (2023). *Is ChatGPT a good translator? Yes, with GPT-4 as the engine*. arXiv preprint arXiv:2301.08745.
- Johansson, R. (2015). *Numerical Python. A Practical Techniques Approach for Industry*. Apress. <https://doi.org/10.1007/978-1-4842-0553-2>
- Khalil, M., & Er, E. (2023). *Will ChatGPT get you caught? Rethinking of plagiarism detection*. arXiv preprint arXiv:2302.04335. <https://doi.org/10.35542/osf.io/fnh48>
- Koceska, N., & Koceski, S. (2014). Financial-Economic Time Series Modelling and Prediction Techniques Review. *Journal of Applied Economics and Business*, 2(4), 28-33.
- Koceska, N., & Koceski, S. (2022). Supporting elderly's independent living with a mobile robot platform. *The Journal of Universal Computer Science-JUCS*, 28(5), 475-498. <https://doi.org/10.3897/jucs.76579>
- Koceska, N., Koceski, S., Beomonte, P.Z., Trajkovik, V., & Garcia, N. (2019). A telemedicine robot system for assisted and independent living. *Sensors*, 19(4), 834. <https://doi.org/10.3390/s19040834>
- Koceska, N., Koceski, S., Lazarova, L.K., Miteva, M., & Zlatanovska, B. (2023). Using ChatGPT for Numerical Solution of First and Second Order Ordinary Differential Equations. In *International Conference of FMNS*. Blagoevgrad, Bulgaria.
- Koceski, S., & Koceska, N. (2016) Evaluation of an assistive telepresence robot for elderly healthcare. *Journal of Medical Systems*, 40, No 5, 121. <https://doi.org/10.1007/s10916-016-0481-x>
- Koceski, S., & Petrevska, B. (2012). Empirical evidence of contribution to e-tourism by application of personalized tourism recommendation system. *Annals of the Alexandru Ioan Cuza University-Economics*, 59(1), 363-374. <https://doi.org/10.2478/v10316-012-0025-6>
- Koceski, S., Panov, S., Koceska, N., Beomonte, P.Z., & Durante, F. (2014). A novel quad harmony search algorithm for grid-based path finding. *International Journal of Advanced Robotic Systems*, 11(9), 144. <https://doi.org/10.5772/58875>
- Kocev, I., Achkoski, J., Bogatinov, D., Koceski, S., Trajkovik, V., Stevanoski, G. et al. (2018). Novel approach for automating medical emergency protocol in military environment. *Technology and Health Care*, 26(2), 249-261. <https://doi.org/10.3233/THC-170852>
- Kocmi, T., & Federmann, C. (2023). Large language models are state-of-the-art evaluators of translation quality. *ArXiv*, 2302.14520.
- Kortemeyer, G. (2023). Could an article-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, 19, 010132. <https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Kumar, V., & Boulanger, D. (2020). Explainable automated essay scoring: Deep learning really has pedagogical value. *Frontiers in Education*, 5. <https://doi.org/10.3389/feduc.2020.572367>
- Kuzman, T, Mozetic, I., & Ljubescic, N. (2023). ChatGPT: Beginning of an end of manual linguistic data annotation? Use case of automatic genre identification. *ArXiv*, abs/2303.03953. <https://doi.org/10.48550/arXiv.2303.03953>
- Lazarova, K.L., Stojkovic, N., Miteva, M., & Stojanova, A. (2021). Application of differential equations in epidemiological models. *Balkan Journal of Applied Mathematics and Informatics*, 4(2), 91-102.
- Lebl, J. (2022). *Notes on Diffy Qs. Differential Equations for Engineers*. Amazon KDP edition.
- Lederman, A., Lederman, R., & Verspoor, K. (2022). Tasks as needs: reframing the paradigm of clinical natural language processing research for real-world decision support. *Journal of the American Medical Informatics Association*, 29(10), 1810-1817. <https://doi.org/10.1093/jamia/ocac121>

- Lehnert, K. (2023). AI insights into theoretical physics and the SwampLand Program: A journey through the cosmos with ChatGPT. *arXiv:2301.08155*.
- Loginova, E.A. (2020). *Application of ordinary differential equations in chemistry*. <https://doi.org/10.18411/lj-07-2020-25>
- Loshkovska, S., & Koceski, S. (2015). *ICT innovations 2015: Emerging technologies for better living*. Springer. <https://doi.org/10.1007/978-3-319-25733-4>
- Megahed, F.M., Chen, Y.J., Ferris, J.A., & Jones-Farmer, L.A. (2023). How generative ai models such as ChatGPT can be (mis) used in spc practice, education, and research? an exploratory study. *Quality Engineering*, 4, 1-29.
- Mishi A.H., Sabari, A.I., Amos, D.A., Egbogu, C.F., Kuje C.A., & Ojosipe, J.O. (2020). Application of Differential Equations in Physics. *Global Scientific Journal*, 8(9), 757-773.
- Momoniati, E., Myers, T.G., Banda, M., & Charpin, J. (2012) Differential equations with applications to industry. *International Journal of Differential Equations*, Article ID 491874. <https://doi.org/10.1155/2012/491874>
- Nagle, R.K., Edward, B.S., & Snider, A.D. (2012). *Fundamentals of Differential Equations*. Pearson Education.
- Nagy, G. (2021). *Ordinary Differential Equations*. Michigan State University Press.
- Nov, O., Singh, N., & Mann, D.M. (2023). Putting ChatGPT's medical advice to the (Turing) test. *JMIR Publications*, 9:e46939. <https://doi.org/10.2196/preprints.46939>
- OpenAi website (n.d.). Available at: <https://openai.com/chatgpt>
- Pardos, Z.A., & Bhandary, S. (2023). Learning gain differences between ChatGPT and human tutor generated algebra hints. *ArXiv*, 2302.06871.
- Polaya, G. (2004). *How to solve it: A new aspect of mathematical method*. Princeton University Press.
- Rangelov, T.V., Dineva, P.S., & Manolis, G.D. (2020). Numerical Solution of Integro-Differential Equations Modelling the Dynamic Behavior of a Nano-Cracked Viscoelastic Half-Plane. *Cybernetics and Information Technologies*, 20(6), 105-115. <https://doi.org/10.2478/cait-2020-0065>
- Shakarian P., Koyalamudi, A., Ngu, N., & Mareedu, L. (2023). An Independent Evaluation of ChatGPT on Mathematical Word Problems (MWP). *ArXiv*, 2302.13814.
- Sobania D., Briesch, M., Hanna, C., & Petke, J. (2023). An analysis of the automatic bug fixing performance of ChatGPT. *ArXiv*, 2301.08653. <https://doi.org/10.1109/APR59189.2023.00012>
- Stamov, T. (2021). Neural networks in engineering design: Robust practical stability analysis. *Cybernetics and Information Technologies*, 21(4), 3-14. <https://doi.org/10.2478/cait-2021-0039>
- Stojanov, D., & Koceski, S. (2014). Topological MRI prostate segmentation method. *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems* (219-225). IEEE. <https://doi.org/10.15439/2014F204>
- Stojanov, D., Mileva, A., & Koceski, S. (2012). A new, space-efficient local pairwise alignment methodology. *Advanced Studies in Biology*, 4(2), 85-93.
- Strang, G. (2022). *Differential Equations and Linear Algebra*. Wellesley, Cambridge Press, MIT.
- Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial Intelligence in Education AIED for Personalised Learning Pathways. *Electronic Journal of e-Learning*, 20(5), 639-653. <https://doi.org/10.34190/ejel.20.5.2597>

- Tenenbaum M., & Pollard, H. (1963). *Ordinary Differential Equations Python. An Elementary Textbook for Students of Mathematics, Engineering, and the Sciences*. Dover Publications.
- Terziyska, M., Todorov, Y., Miteva, D., Doneva, M., Dyankova, S., Metodieva, P. et al. (2020). Regularized Neural Network for Prediction of the Dose in Gamma Irradiated Milk Products. *Cybernetics and Information Technologies*, 20(2), 141-151. <https://doi.org/10.2478/cait-2020-0022>
- Trajkovik, V., Gjorgievska, E.V., Koceski, S., & Kulev, I. (2014). General assisted living system architecture model. *International Conference on Mobile Networks and Management* (329-343). Springer, Cham. [https://doi.org/10.1007/978-3-319-16292-8\\_24](https://doi.org/10.1007/978-3-319-16292-8_24)
- Tu R., Ma, C., & Zhang, C. (2023). Causal-discovery performance of ChatGPT in the context of neuropathic pain diagnosis. *ArXiv*, 2301.13819.
- Wang, J., Liang, Y., Meng, F., Shi, H., Li, Z., Xu, J. et al. (2023). Is ChatGPT a good nlg evaluator? A preliminary study. *ArXiv*, 2303.04048. <https://doi.org/10.18653/v1/2023.newsum-1.1>
- West C.G. (2023). AI and the FCI: Can ChatGPT project an understanding of introductory physics? *ArXiv*, 2303.01067.

Published by OmniaScience ([www.omniascience.com](http://www.omniascience.com))

Journal of Technology and Science Education, 2025 ([www.jotse.org](http://www.jotse.org))



Article's contents are provided on an Attribution-Non Commercial 4.0 Creative commons International License. Readers are allowed to copy, distribute and communicate article's contents, provided the author's and JOTSE journal's names are included. It must not be used for commercial purposes. To see the complete licence contents, please visit <https://creativecommons.org/licenses/by-nc/4.0/>.