Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 20 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i20.42979

PAPER

Learning Mathematics with Large Language Models: A Comparative Study with Computer Algebra Systems and Other Tools

Nikolaos Matzakos[1](✉), **Spyridon Doukakis**[2], **Maria Moundridou**[1]

[1]School of Pedagogical and Technological Education, Athens, Greece

[2]Ionian University, Corfu, Greece

nikmatz@aspete.gr

ABSTRACT

Artificial intelligence (AI) has permeated all human activities, bringing about significant changes and creating new scientific and ethical challenges. The field of education could not be an exception to this development. OpenAI's unveiling of ChatGPT, their large language model (LLM), has sparked significant interest in the potential applications of this technology in education. This paper aims to contribute to the ongoing discussion on the role of AI in education and its potential implications for the future of learning by exploring how LLMs could be utilized in the teaching of mathematics in higher education and how they compare to the currently widely used computer algebra systems (CAS) and other mathematical tools. It argues that these innovative tools have the potential to provide functional and pedagogical opportunities that may influence changes in curriculum and assessment approaches.

KEYWORDS

large language models (LLMs), ChatGPT, bard experiment, computer algebra systems (CAS), maxima, computational knowledge engines, Wolfram|Alpha, mathematical education, higher education

1 INTRODUCTION

More than half a century has passed since the beginning of the use of digital technologies in the teaching and learning of mathematics. In the early 1970s, the attempt to introduce simple calculators into higher education led to a debate about their value and appropriateness, while their use would inevitably change the context of the mathematics classroom. Etlinger's observation in 1974 highlighted the dual potential of technology, encompassing functional utility in generating responses and pedagogical significance in augmenting the learning process [1]. Subsequently, technology introduced a plethora of mathematical tools to educators, including software specifically designed for calculations, graphing, and precise diagrammatic

Matzakos, N., Doukakis, S., Moundridou, M. (2023). Learning Mathematics with Large Language Models: A Comparative Study with Computer Algebra Systems and Other Tools. *International Journal of Emerging Technologies in Learning (iJET)*, 18(20), pp. 51–71. https://doi.org/10.3991/ijet.v18i20.42979

Article submitted 2023-06-10. Revision uploaded 2023-08-03. Final acceptance 2023-08-03.

representation. These tools hold promising prospects for facilitating advantageous shifts in teaching and learning methodologies, leading to the transformation of the teacher's role and the redefinition of the tasks given to the students.

In the following decades, research on the one hand and advancements in software development on the other brought to the fore new digital technologies that could be integrated into the mathematics classroom. These technologies include computer algebra systems (CASs), dynamic geometry software, and statistical packages. This development has sparked rich discussions about the roles of these tools and their optimal utilization in educational settings. Noteworthy is the work of Pierce and Stacey in [2], who identified the opportunities that arise from the integration of mathematical software into the educational process. However, even in the present era, research such as [3] highlights that the utilization of mathematical software in higher education remains limited, primarily due to the absence of a comprehensive framework associated with the integration of technology in mathematics education.

While the development of a precise framework for integrating such technologies into mathematics education is still ongoing, the emergence of artificial intelligence systems, particularly large language models (LLMs), brings forth yet another digital technology to be considered. The recent widespread adoption of LLMs, exemplified by models such as ChatGPT, and their unrestricted usage have already initiated numerous discussions and debates regarding the positive and negative perspectives [4]. Thus, the following question arises: Since open access to these models cannot limit or prohibit their use, how can they be effectively incorporated into the realm of mathematical education to enhance teaching and learning? Therefore, there is a need to identify the teaching and learning opportunities offered by LLMs and explore their influence on the curriculum and classroom dynamics, particularly in higher education settings for science and engineering studies. Mapping the opportunities offered by the use of LLMs can be a starting point for further research exploring the effectiveness and impact of the integration of LLMs on students' learning outcomes. In addition, there is a prospect of highlighting specific areas where educational technology can capitalize on the rich possibilities offered by LLMs, including tailored learning experiences, real-time feedback, and the enhancement of digital skills.

The next section of this paper provides a concise overview of CASs and other mathematical problem-solving tools that are currently utilized as teaching and learning aids in the mathematics classroom. This is followed by an exploration of the emerging artificial intelligence tools that hold promise in the same context. Subsequently, in order to compare LLMs to these other tools and illustrate their capabilities and limitations, four practical problem-solving cases from calculus, linear algebra, and numerical analysis—addressed with all three different tools—are presented. It is then attempted to map the new opportunities arising from the integration of LLMs in mathematics education and discuss their potential to enhance the teaching and learning of mathematics.

2 THEORETICAL BACKGROUND

2.1 CAS in higher education mathematics

In recent decades, the significant progress that has been made in the development of numerical and symbolic mathematical software, including CASs, has led to transformative shifts in teaching methods and problem-solving approaches in higher education mathematics [5] [6]. A CAS is a specialized software designed to perform symbolic mathematical calculations, such as algebraic manipulations, calculus, and matrix operations. The first CAS was created at MIT in the late 1970s and was known as Mascyma. Today, there is a wide range of such programs available, both as free/open-source software (e.g., Axiom, Erable, Fermat, Maxima, OpenAxiom, Reduce, etc.) and proprietary software (e.g., Derive, Mathematica, Mathcad, and Maple). CASs have been widely utilized in various fields such as engineering, finance, physics, and chemistry, enabling researchers and professionals to quickly perform complex calculations. As highlighted in the literature [7, 8], users should verify the results produced by a CAS by cross-checking them with multiple CASs and/or by using digital mathematical libraries (DML).

Since the early 1980s, efforts have been made to integrate CASs into education, prompting investigations into their methods and suitability as educational tools. Initial research yielded encouraging results concerning their use [9] [10]. In addition, researchers referred to the opportunities that CASs offer both teachers as a means to transform their teaching approach and to students who have in their quiver a tool for calculating, modeling, and understanding mathematical concepts [11–13].

Furthermore, other research highlights a more nuanced perspective on the potential impact of CASs on education [14]. These investigations study the impact of CAS on learning outcomes and attempt to identify the advantages and challenges associated with their integration into mathematics teaching. The instrumental approach serves as a conceptual framework that supports the research and integration of CAS in education [15, 16]. According to this perspective, CAS are considered tools with which students can acquire mathematical knowledge, and at the same time, teachers have an essential role in orchestrating their use to facilitate student learning, turning CAS into an "instrument". The use of CASs is distinguished based on their pragmatic and cognitive value, with students using them primarily for reasons of efficiency and answer validation. However, there is a growing tendency to utilize CASs scientifically, i.e., delegate the calculations to the tool and allow students to focus on exploring mathematical concepts and interpreting the results obtained from the tool [14]. This shift in approach highlights the importance of promoting deeper mathematical understanding and conceptual reasoning.

Finally, advances in CASs' technology, including the integration of AI, have opened up new possibilities for users [17], enabling the automatic discovery and proof of geometric properties and providing tutorials for solving mathematical problems. As a result, CASs are a critical component of the toolbox for modern mathematicians and engineers.

2.2 Online problem-solving tools in higher education mathematics

There are numerous online platforms available that facilitate the solving of scientific problems by providing not only the solutions but also the step-by-step process leading to those solutions. Examples of such platforms include Wolfram|Alpha, Mathway, Symbolab, Cymath, and others. These platforms offer the capability to input math problems in natural language or unstructured formats, and they employ CASs to generate relevant answers. In recent times, numerous researchers have been exploring and discussing the advantages and drawbacks associated with the utilization of such tools in higher education [18–21]. One distinctive feature of these platforms is their computability, as they employ computational algorithms to process and solve problems.

Wolfram|Alpha, the tool selected as a representative of this category in this study, stands out as a broader knowledge engine that extends beyond mathematics.

It covers a wide range of domains, including science, finance, linguistics, and more. Wolfram|Alpha is a computational knowledge engine that automatically provides answers to questions while also delivering extensive analyses and graphs through its curated data and sophisticated algorithms [22]. Users can input questions in natural language or use the available tools to perform complex calculations and data analyses. Wolfram|Alpha is a reliable tool for scientists, engineers, students, and anyone seeking immediate access to scientific knowledge. It is built with the Wolfram Language [23], which is a symbolic language deliberately designed with the breadth and unity needed to develop powerful programs quickly. To function, it converts natural language, where possible, into Wolfram Language and then carries out the computations according to its algorithms. As it can be noticed, Wolfram|Alpha shares common features and functions with both CASs and large language models.

Wolfram|Alpha has been utilized in the context of mathematical education, such as calculus and linear algebra [24–26]. It offers several advantages over CASs, including its natural language interface, platform independence, and the fact that it functions as a computational knowledge engine aiming to answer broader scientific questions. The capability of Wolfram|Alpha to process calculations expressed in natural language has been a long-standing goal of the research community, which has pursued it through various projects often involving chatbots. One notable example is the Sofia AI Project, which features a calculus chatbot capable of receiving questions in natural language and returning the corresponding calculations [27] [28]. The aforementioned research has played a significant role in paving the way and igniting interest in exploring the integration of LLMs in mathematics education.

2.3 LLMs in higher education mathematics

Following the introduction of the ChatGPT, LLMs gained significant attention, especially regarding their potential impact on almost every kind of human activity. These advanced AI models are designed to process and generate human-like text by utilizing extensive amounts of pre-existing language data. There are other such models, including the Bard experiment, bidirectional encoder representations from transformers (BERT), generative pre-trained transformer-2 (GPT-2), extreme multilingual language understanding (XLNet), robustly optimized BERT approach (RoBERTa), text-to-text transfer transformer (T5), and pathways language model (PaLM) [29–32].

Education is a field of human activity that constantly explores and strives to integrate new technological tools as soon as they emerge. The utilization of LLMs in education is being increasingly discussed due to their potential impact on the field. In mathematical education, efforts are being made to discover ways to utilize this new technology, particularly for mathematical computations. An exemplary precursor to such an artificial intelligence system employing natural language is the Sofia AI project, which was mentioned in the previous section [27]. A recent study [33] presented an LLM, called Minerva, based on the language model PaLM. This model is geared towards problems in mathematics, physics, and engineering, and its performance has been assessed. They found that the model could correctly answer almost one-third of the two hundred undergraduate-level problems it was fed. Researchers [34] investigated the mathematical capabilities of ChatGPT and measured its performance relative to Minerva. The mathematical problems posed to the system are not elementary; thus, their resolution poses a significant challenge to AI systems. They concluded that, contrary to formal mathematics, where large databases of formal proofs exist (e.g., the Lean's Library of Formal Mathematical Proofs), the datasets of mathematics in natural language used to evaluate language models only cover elementary mathematics. The researchers infer that on these specific datasets, the mathematical abilities of ChatGPT are comparatively lower than those of a graduate student. Furthermore, while the system seems to understand the question, it does not always succeed in returning the correct solution [35, 36].

3 SOLVING MATHEMATICAL PROBLEMS USING CASS, ONLINE PROBLEM-SOLVING TOOLS AND LLMS

The heightened interest in leveraging LLMs in education necessitates an exploration of the extent to which they can substitute or supplement CAS for solving mathematical problems, especially within the context of mathematics education. For this purpose, problems from calculus, linear algebra, and mathematical analysis will be presented. These problems will be solved using the aforementioned tools in order to derive a comparative evaluation of their capabilities. Specifically, the selected tools for the study were Maxima (CAS), the computational knowledge engine Wolfram|Alpha as a representative of the online problem-solving tools, and ChatGPT-4 as an LLM. It should be noticed that in the case of the LLMs, the Bard experiment was also tested, and the results it generated were similar to those of ChatGPT. However, it was chosen to report only the interactions with ChatGPT, as it is currently the more widely used and easily accessible of the two.

It emerges that in relatively simple operations and symbolic calculations such as, for example, arithmetic operations, basic integrals, and equations, all types of tools yield reliable results. Hence, with ChatGPT and Wolfram|Alpha users are allowed to solve a simple problem stated in natural language, without having to know the CAS commands. For instance, to compute $\int xe^x dx$ in Maxima, a user needs to know the `'integrate'` command and its syntax: `integrate(x*exp(x),x)`. On the other hand, in Wolfram|Alpha and ChatGPT, the user can pose the question "*What is the integral of xexp(x)?*" and receive the correct answer, along with the calculation method.

However, in slightly more complex calculations, such as computing high powers of integers or more intricate integrals, ChatGPT seems to fall short [23] [34], unlike Maxima and Wolfram|Alpha which return correct results. Yet, in ChatGPT, the user can point out the erroneous result and guide the LLM to generate the correct answer.

The following subsections present the attempts to solve four typical problems from the curriculum of an undergraduate mathematics course using the aforementioned tools.

3.1 Problem 1: The inverse of the square matrix

As an example, we try to compute the inverse of the square matrix

$$A = \begin{bmatrix} 1 & -1 & 1 \\ -2 & 1 & -3 \\ 1 & -2 & 1 \end{bmatrix}$$

A. With a CAS *(Maxima)* the user needs to enter the following commands:

```
(%i) A:matrix([1,-1,1],[-2,1,-3],[1,-2,1]);
(%i) invert(A);
```

Maxima returns $\begin{bmatrix} 5 & 1 & -2 \\ 1 & 0 & -1 \\ -3 & -1 & 1 \end{bmatrix}$ which is indeed the correct answer.

B. With *Wolfram|Alpha* the user requests:

Compute the invert matrix of [[1, –1, 1], [–2, 1, –3], [1, –2, 1]]

Wolfram|Alpha returns the correct result. It also provides the step-by-step solution using the Gaussian elimination method.

C. With *ChatGPT* the user asks:

Compute the invert matrix of [[1, –1, 1], [–2, 1, –3], [1, –2, 1]]

ChatGPT replies as follows:
To find the inverse of the matrix, we can use the following formula:
*A^ – 1 = (1/det(A)) * adj(A)*
where det(A) is the determinant of A, and adj(A) is the adjugate matrix of A.
First, let's compute the determinant of the matrix:
det(A) = 1(11 – (–2)(–2)) – (–1)(–2(–3)) + 1*(–2*(–2) – (–1)*(–3)) = 1 – 6 + 2 = –3*
Next, we need to compute the adjugate matrix of A, which is the transpose of the matrix of cofactors of A. The matrix of cofactors is obtained by taking the determinants of the 2 × 2 matrices formed by deleting each element of A, and multiplying them by the corresponding sign (–1)^(i + j), where i and j are the row and column indices of the element.
So, the matrix of cofactors is:
C = [[1, –1, 1], [–2, –1, 2], [1, –1, 1]]
Taking the transpose of C, we get:
adj(A) = [[1, –2, 1], [–1, –1, –1], [1, 2, 1]]
Finally, we can compute the inverse of A using the formula mentioned above:
*A^ – 1 = (1/det(A)) * adj(A) = (–1/3) * [[1, –2, 1], [–1, –1, –1], [1, 2, 1]]*
= [[–1/3, 2/3, –1/3], [1/3, 1/3, 1/3], [–1/3, –2/3, –1/3]]
Therefore, the inverse of the matrix [[1, –1, 1], [–2, 1, –3], [1, –2, 1]] is [[–1/3, 2/3, –1/3], [1/3, 1/3, 1/3], [–1/3, –2/3, –1/3]].
The system describes all steps correctly based on the appropriate methods, but the final result it produces is numerically wrong. If the user points out that the answer is wrong, *ChatGPT* regenerates its response using either the same or a different method but still returns a numerically wrong final answer.
In order to overcome this pitfall, a user may ask *ChatGPT* to solve the problem using a CAS for the calculations:

Compute the invert matrix of [[1, –1, 1], [–2, 1, –3], [1, –2, 1]].
Use Maxima for your calculations.

ChatGPT in this case does not describe the way to solve the problem but provides the Maxima commands that should be used without executing them. The user themselves may use these commands and solve the problem in Maxima.

3.2 Problem 2: Find the minimum-maximum value of a two-variable function

As an example, we try to find the min-max of the function $f(x, y) = 1-(x^2 + y^2)$

A. With a CAS *(Maxima)* the user needs to enter the following commands:

```
(%i) f(x,y):=1-(x^2+y^2);
(%i) dfx:diff(f(x,y),x);
(%i) dfy:diff(f(x,y),y);
(%i) sol:solve([dfx=0,dfy=0],[x,y]);
(%i) h1:determinant(hessian(f(x,y),[x]));
(%i) h2:determinant(hessian(f(x,y),[x,y]));
(%i) subst([sol[1]],h2);
(%i) subst([sol[1]],h1);
```

With these commands though, Maxima performs the calculations and returns the numerical results. To find the solution of the problem, the user should apply the relevant mathematical theory to see that the maximal value is on the point (0, 0) and the maximal value of the function is 1.

B. With *Wolfram|Alpha* the user requests:

Find the min max value of 1 – (x^2 + y^2);

Wolfram|Alpha performs several calculations and returns the results. Among those, the user can find the global maximum value 1 at the point (0, 0).

C. With *ChatGPT* the user asks:

Find the min max values of 1 – (x^2 + y^2);

ChatGPT replies as follows:
To find the min-max values of the two-variable function f(x, y) = 1 – (x^2 + y^2), we need to find its critical points, which occur when the partial derivatives with respect to x and y are equal to zero or are undefined.
Let's compute the first partial derivatives of the function:
∂f/∂x = –2x ∂f/∂y = –2y
Now, let's find the critical points by setting the partial derivatives equal to zero:
–2x = 0 => x = 0 –2y = 0 => y = 0
The only critical point is at (x, y) = (0, 0).
To classify this critical point, we'll compute the second partial derivatives:
$∂^2f/∂x^2 = –2$ $∂^2f/∂y^2 = –2$ $∂^2f/∂x∂y = 0$
Now, we'll compute the discriminant, D, using the following formula:
$D = (∂^2f/∂x^2)(∂^2f/∂y^2) – (∂^2f/∂x∂y)^2$
$D = (–2)(–2) – 0^2 = 4$
Since D > 0 and $∂^2f/∂x^2 = –2$ is negative, the critical point (0, 0) corresponds to a local maximum.
Now let's find the function value at this local maximum:
f(0, 0) = 1 – (0^2 + 0^2) = 1
Thus, the local maximum value of the function f(x, y) = 1 – (x^2 + y^2) is 1 at the point (0, 0).
Since the function represents a paraboloid that opens downward, there is no minimum value for the function. As (x, y) moves away from the origin, the function will continue to decrease, approaching negative infinity. Therefore, the function has a local maximum value of 1 at the point (0, 0) but does not have a minimum value.
ChatGPT in this case returns the correct result along with all the steps and calculations it followed to reach the solution.

3.3 Problem 3: Double integrals over general regions

As an example, we try to calculate the double integral of the function sin(x)/x over the triangular region bounded $x = 0$, $y = x$, and $x = 1$.

A. With a CAS: If using *Maxima* for this calculation, one should manually express the region of integration by drawing it as illustrated in Figure 1. This region is $= \{(x, y) \in \mathbb{R}^2 \mid 0 \le x \le 1 \kappa\alpha\iota\, 0 \le y \le x\}$. Then, the user calculates the double integral as integrated integral $\iint_R f(x, y)dA = \int_0^1 \left[\int_0^x \frac{\sin x}{x} dy \right] dx$.

```
(%i) draw2d(grid=true, xaxis=true,
yaxis=true, xrange=[0,1.5],
yrange=[0,2,color=black, fill_color=grey,
filled_func=x, explicit (0,x,0,1),
filled_func=false, explicit(x,x,0,1.5),
point_size=1, point_type=filled_circle,
points([[1,1]]), label([“(1,1)”,1.2,1]));
```

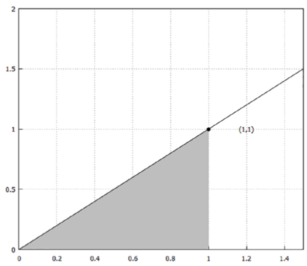

Fig. 1. Region of the integral

Finally, by applying the *Maxima* commands, the user obtains the correct result.

```
(%i) f(x,y):=sin(x)/x;
(%i) integrate(integrate(f(x,y),y,0,x),x,0,1);
(%o) 1-cos(1)
```

B. With *Wolfram|Alpha* the user requests:

Calculate the double integral of the function sin(x)/x over the triangular region bounded x = 0, y = x and x = 1

Wolfram|Alpha does not return any answer and requests that the user rephrases the problem.

C. With *ChatGPT* the user asks:

Calculate the double integral of the function sin(x)/x over the triangular region bounded from x = 0, y = x and x = 1

ChatGPT replies as follows:
Since the region is bounded by x = 0, y = x, and x = 1, the limits for x are 0 to 1, and for y, it will vary from 0 to x. The double integral can be set up as follows:
∬[sin(x)/x] dy dx
Now let's set the limits:
∫(x = 0 to x = 1) ∫(y = 0 to y = x) [sin(x)/x] dy dx
Notice that there is no 'y' term in the function, so we can treat the inner integral as a constant with respect to y. Thus, the integral becomes:
*∫(x = 0 to x = 1) [sin(x)/x * ∫(y = 0 to y = x) dy] dx*
Integrating the inner integral with respect to y:
∫(y = 0 to y = x) dy = [y] (from 0 to x) = x
Now, the double integral simplifies to:
*∫(x = 0 to x = 1) [sin(x)/x * x] dx*
The x terms cancel out, and the integral becomes:
∫(x = 0 to x = 1) sin(x) dx
Now, integrate sin(x) with respect to x:
∫ sin(x) dx = –cos(x)

Evaluate the integral from 0 to 1:
[–cos(x)] (from 0 to 1) = –cos(1) – (–cos(0)) = –cos(1) + 1
So, the value of the double integral over the given region is approximately – cos(1) + 1
ChatGPT in this case returns the correct result along with all the steps and calculations it followed to reach the solution.

3.4 Problem 4: Least squares method

As an example, we try to find the linear least squares model for the data [0.2, 0.52], [0.5, 0.9], [0.6, 1], [0.9, 1.3],[1, 1.5], [1.1, 1.7]

A. With a CAS (*Maxima*):

```
(%i) A: matrix ([0.2, 0.52], [0.5, 0.9], [0.6, 1],
      [0.9,1.3],[1,1.5],[1.1,1.7]);
(%i) load (lsquares);
(%i) ls_model:lsquares_estimates(A, [x,y], y=a*x+b,
      [a,b]);
(%i) float(ls_model);
```

Maxima returns [[a = 1.248725212464589, b = 0.258413597733711]] which is indeed the correct answer.

B. With *Wolfram|Alpha* the user requests:

Find the linear least squares model for the data [0.2, 0.52], [0.5, 0.9], [0.6, 1], [0.9,1.3], [1,1.5], [1.1,1.7]

Wolfram|Alpha returns the correct result. But it does not provide the step-by-step solution.

C. With *ChatGPT* the user asks:

Find the linear least squares model for the data [0.2, 0.52], [0.5, 0.9], [0.6, 1], [0.9, 1.3],[1, 1.5], [1.1, 1.7]

ChatGPT replies as follows:
To find the least squares model for the given data, we need to fit a linear equation of the form y = mx + b to the data points. The least squares method minimizes the sum of the squared residuals between the predicted y-values and the actual y-values.
Let's calculate the least squares model for the given data:
Data points: [0.2, 0.52] [0.5, 0.9] [0.6, 1] [0.9, 1.3] [1, 1.5] [1.1, 1.7]
Step 1: Calculate the mean of x and y values. Sum of x-values: 0.2 + 0.5 + 0.6 + 0.9 + 1 + 1.1 = 4.3 Sum of y-values: 0.52 + 0.9 + 1 + 1.3 + 1.5 + 1.7 = 6.92
Mean of x-values: 4.3 / 6 = 0.7167 (rounded to 4 decimal places) Mean of y-values: 6.92 / 6 = 1.1533 (rounded to 4 decimal places)
Step 2: Calculate the deviations of x and y values from their means. Deviation of x-values: [0.2 – 0.7167, 0.5 – 0.7167, 0.6 – 0.7167, 0.9 – 0.7167, 1 – 0.7167, 1.1 – 0.7167] = [–0.5167, –0.2167, –0.1167, 0.1833, 0.2833, 0.3833]
Deviation of y-values: [0.52 – 1.1533, 0.9 – 1.1533, 1 – 1.1533, 1.3 – 1.1533, 1.5 –1.1533, 1.7 – 1.1533] = [–0.6333, –0.2533, – 0.1533, 0.1467, 0.3467, 0.5467]
*Step 3: Calculate the sum of the squared deviations of x-values and the product of deviations of x and y values. Sum of squared deviations of x-values: (–0.5167)^2 + (–0.2167)^2 + (–0.1167)^2 + (0.1833)^2 + (0.2833)^2 + (0.3833)^2 = 0.425 Product of deviations of x and y values: (–0.5167) * (–0.6333) + (–0.2167) * (–0.2533)*

*+ (–0.1167) * (–0.1533) + (0.1833) * (0.1467) + (0.2833) * (0.3467) + (0.3833) * (0.5467) = 0.6249*
Step 4: Calculate the slope (m) and y-intercept (b) of the least squares line. m = product of deviations of x and y values / sum of squared deviations of x-values = 0.6249 / 0.425 = 1.4694 (rounded to 4 decimal places)
*b = mean of y-values – (m * mean of x-values) = 1.1533 – (1.4694 * 0.7167) = 0.1012 (rounded to 4 decimal places)*
Therefore, the least squares model for the given data is: y = 1.4694x + 0.1012
The system describes all steps correctly based on the appropriate methods, but the final result it produces is numerically wrong.
Upon observing these four fundamental mathematical problems, various issues arise regarding the reliability of the results generated by LLMs. The following section presents both the advantages and disadvantages of the aforementioned approaches and introduces strategies that aim to enhance the efficiency and validity of computations while also maximizing the pedagogical benefits associated with the use of large language models.

4 FUNCTIONAL AND PEDAGOGICAL OPPORTUNITIES OF LLMS

In the last few decades, CASs have been widely used in mathematics education, serving as valuable tools that offer several opportunities for students as well as educators [2] [3] [9] [11] [13] [15]. Recent technological advancements, notably in the realm of AI, have extended these opportunities further and even introduced new ones. In particular, LLMs have emerged as a significant development in the field of AI, revolutionizing the way students interact with mathematical concepts and presenting new opportunities and challenges in mathematics education.

Although it is acknowledged in Section 3 that LLMs may not consistently meet the desired level of reliability, they still demonstrate potential functional and pedagogical opportunities. Furthermore, the combination of LLMs with CASs and DMLs shows promise in addressing specific functional concerns. In the upcoming section, we will delve into strategies for effectively and reliably integrating LLMs into mathematics education and research.

4.1 Functional opportunities

The functional limitations of LLMs in performing reliable mathematical calculations can potentially be addressed by integrating them with complementary tools. Figure 2 illustrates an approach to utilizing LLMs, where users present mathematical problems in natural language by posing one or more questions. The solution generated by the LLM is then reviewed and refined by the user. To ensure the accuracy of the calculations, one possible method is to involve a CAS by querying the LLM to generate the corresponding commands. The calculations can then be independently verified by running these commands in the CAS.

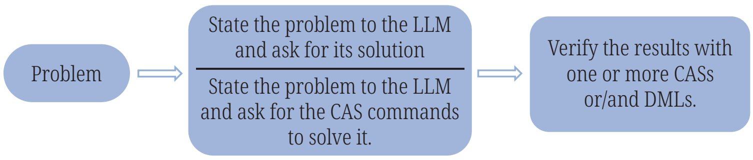

Fig. 2. An approach to utilizing LLMs to solve a mathematical problem

Another way to overcome the possibly unreliable ChatGPT results is to utilize the recently available Wolfram Plugin. As per Wolfram in [23] [37], a way to harness the shared natural language interface between Wolfram|Alpha and LLMs, such as ChatGPT, is to enable ChatGPT to interact with Wolfram|Alpha in a human-like manner. This involves converting the natural language input from ChatGPT into a precise, symbolic computational language, allowing Wolfram|Alpha to effectively utilize its computational knowledge. The interface for this collaboration can be achieved using OpenAI Plugins [38]. By activating the specific Wolfram Plugin, as shown in Figure 3, ChatGPT will generate the correct answer for problem 1 in Section 3.

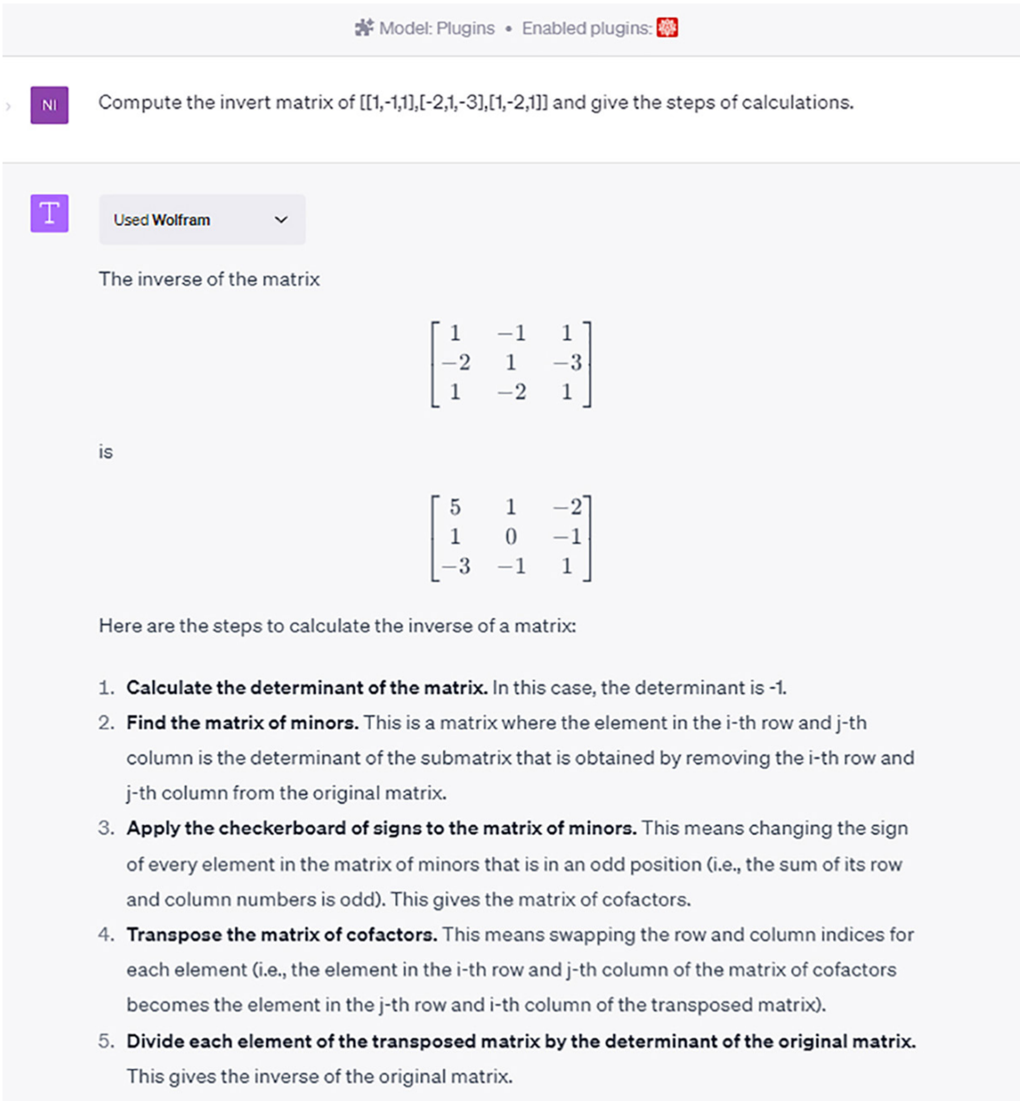

Fig. 3. Posing problem 1 to ChatGPT using Wolfram plugin

While CASs serve as useful tools for performing calculations, they often fall short both in providing the necessary guidance on problem-solving steps and in assisting students understand the underlying concepts. On the contrary, the new LLM tools empower their users to actively engage in interactive communication by expressing mathematical problems in natural language. Moreover, users can pose follow-up questions in order to get more elaborate explanations and insights into the solution process. As a result, LLMs offer explanations for the underlying concepts, bridging the gap between conceptual understanding and step-by-step problem-solving. Consequently, LLMs appear to offer valuable advantages in both dimensions crucial for addressing mathematical problems: comprehending the underlying concepts and employing the appropriate mathematical calculations to find solutions.

The aforementioned approach to the functional utilization of LLMs in mathematical problem-solving allows for the exploration of their potential implications in mathematics education. Similarly to [2], there is evidence that the adoption of these innovative tools will lead to substantial transformations, not only in the instructional content but also in the assessment methods used to evaluate the learning outcomes.

Table 1 presents a concise overview of the primary capabilities and limitations of CASs, online problem-solving tools, and LLMs, as inferred from their application in solving the sample mathematical problems outlined in Section 3.

Table 1. Functional capabilities and limitations of CAS, Online problem-solving tools and LLMs

	Capabilities	Limitations
Computer Algebra Systems (e.g., Maxima, Maple)	– Producing more reliable results. – Generating 2D and 3D graphical representations. – Providing a functional environment with the ability to save calculations in worksheets.	– They do not provide information on the steps required to solve each kind of problem. – Users must be familiar with its commands.
Online problem-solving tools (e.g., Wolfram|Alpha, Mathway)	– Processing questions in natural language. – No need for explicit commands. – Producing more reliable results. – Displaying intermediate results. – Generating 2D and 3D graphical representations. – Integration with LLMs.	– They cannot fully understand questions in natural language. – Lack of a worksheet-type working environment.
Large Language Models (e.g., ChatGPT, Bard)	– Processing questions in natural language. – No need for explicit commands. – Displaying intermediate results along with the solution method. – Providing commands for execution in a CAS to enhance reliability in calculations. – Ability to connect with Wolfram|Alpha for reliable computations.	– Unreliable numerical calculations – Lack of a worksheet-type working environment. – Graph generation is not yet supported.

The following section investigates the pedagogical prospects that may arise from employing LLMs, with the objective of assessing their alignment with the opportunities outlined in the pedagogical map for mathematics analysis software developed by Pierce and Stacey [2].

4.2 Pedagogical opportunities

In the context of mathematical education for students in engineering and the sciences, most academic institutions have embraced the integration of digital technologies, particularly CASs, in both instructional and learning processes.

The emergence of this innovative technological possibility prompted a reassessment of the pedagogical framework in mathematics education. Pierce and Stacey [2] investigated the pedagogical opportunities for educators and learners stemming from the utilization of mathematical software and provided insight for further research [3, 39–41]. These pedagogical opportunities were structured at three distinct levels. Firstly, there are opportunities for enhancing teaching and learning activities, encompassing various tasks that can be accomplished through the use of the software. Secondly, the software creates opportunities for innovative teaching and learning approaches, allowing educators to explore new instructional methods and strategies. Lastly, the subject matter of mathematics itself is positively influenced, with the software enabling deeper exploration and understanding of mathematical concepts. Drawing upon their work, an endeavor will be undertaken to discuss the potential fit of LLMs within this context. Figure 4 presents a pedagogical map for the use of LLMs in mathematics education that will be analyzed in the subsequent paragraphs.

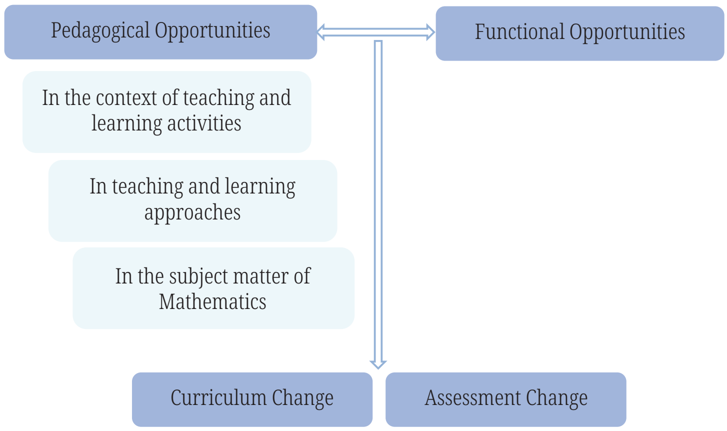

Fig. 4. LLMs' potential impact on pedagogy, curriculum, and assessment

In the context of teaching and learning activities. The cultivation of pen-and-paper skills is reinforced as students are afforded the opportunity to observe and analyze the responses generated by LLMs. These responses not only offer students an answer to a mathematical problem but also provide a detailed description of the steps involved in reaching this solution. However, as evidenced, LLMs occasionally produce incorrect computations. This flaw creates a pedagogical opportunity in which students engage in critical evaluation of the solutions by verifying them either by hand or by using external resources such as CAS or Wolfram|Alpha. Simultaneously, students are allowed to engage in interactive discourse with LLMs, enabling them to pose multiple inquiries and seek elucidations regarding the presented outcomes.

Learning activities that draw on real-world contexts are of great pedagogical value, as they can motivate and engage students in mathematical thinking while demonstrating the real uses of mathematics [42]. When it comes to real-world problems, LLMs may be useful in various ways. They can accept a real problem in natural language and return the steps of its solution. They can also process real data swiftly and provide results (note that the accuracy of the computations must be ensured, as mentioned in the previous paragraph). Moreover, LLMs have the ability to generate real-world problems in natural language as well as create synthetic datasets to serve as test cases [43].

Large language models can be used to observe the effects of parameters and discover patterns. Thus, they can be leveraged to find the solution for a problem for different values of one or more parameters, pose "what if..." questions, and observe regularities. However, in the current version of LLMs, the capability to explore regularity and variation graphically, as would be possible with dynamic geometry software, is not available. Despite this limitation, LLMs can still be utilized for simulation purposes as they can generate the commands for implementing the simulation in a CAS.

Large language models, within the context of problem-solving, interact with the student textually, numerically, and symbolically. This ability for multiple representations is significant, as it is believed to enhance student comprehension [44] [45].

In teaching and learning approaches. The use of LLMs in the teaching and learning process allows for changes in social dynamics and the didactic contract, as educators are no longer the sole authority and students gain greater control over their own learning. In this way, educators may assume a facilitator role and guide students in effectively utilizing LLMs for purposes such as research and analysis. Learners, in turn, may actively engage with LLMs to explore, collaborate, and discover knowledge. Moreover, LLMs provide instant information and expand the range of possible resources available to students and educators, resulting in a more knowledge-rich learning environment in which both are allowed to explore various perspectives and engage in deeper discussions. In addition, differentiated learning is enhanced as the LLMs response is developed based on student questions. In this way, LLMs provide tailored content and supplementary explanations that align with the specific dialogue established with different students or student groups. The need to assess the responses offered by LLMs presents educators with an opportunity to incorporate cooperative and inquiry approaches into the educational process, thus enhancing students' collaborative and critical thinking skills. Furthermore, the increased accessibility of information to students facilitated by LLMs encourages educators to consider redefining their assessment methods. This shift involves evaluating skills and abilities such as critical thinking, creativity, collaboration, and problem-solving rather than solely concentrating on the memorization and testing of declarative knowledge.

The aforementioned factors play a crucial role in reshaping the didactic contract, emphasizing the need to address ethical considerations related to the utilization of LLMs. These considerations include source evaluation, plagiarism prevention, and compliance with copyright regulations. Moreover, the redesign of the learning environment nurtures a new dynamic that enhances learner autonomy and transforms the teacher's role in delivering instruction and implementing assessment methods so as to align with the dynamic educational landscape enabled by the integration of large language models.

In the subject matter of Mathematics. Similarly to the transformative effect that has been observed from the utilization of mathematical software [2], the integration of LLMs in mathematics education profoundly impacts the content taught, the level of comprehension demanded for mathematical concepts, and the interrelationships among these concepts.

The occurrence of errors during LLM calculations creates an environment within the classroom that encourages discussions among students, between students and instructors, as well as between students and the LLM system. These dialogues serve to enhance the understanding of mathematical concepts. Furthermore, LLMs provide an opportunity for educators to emphasize the acquisition of conceptual understanding rather than mere procedural skills. This approach cultivates students' mathematical thinking and facilitates a deeper comprehension of the subject matter. For instance, in a scenario involving an optimal linear model problem, students might initially rely on direct formula-based calculations. However, through engagement with LLMs, students gain access to the process of proving and constructing these formulas, thereby encountering additional concepts such as partial derivatives and multivariable function optimization. As a result, class discussions centered on these concepts can arise, effectively reshaping the content of mathematical education. Each student can seek further elucidation on these concepts from the LLMs, enabling individual knowledge enrichment and stimulating productive classroom discussions that promote a deeper understanding of mathematical concepts.

Simultaneously, LLMs provide educators with an opportunity to underscore the importance of mastering concepts and applying them practically, moving beyond mere skill acquisition. Engaging students in activities where they pose mathematical problems to an LLM and examine the provided solutions and explanations allows them to concentrate on understanding the related concepts, while skill development can be addressed thereafter.

By integrating LLMs into the teaching process, educators can explore alternative pathways and educational methodologies. This may involve the instructor asking the LLM to create real-world problems or provide solutions to these problems based on actual or generated data. Such an approach offers students a comprehensive and meaningful learning experience, allowing them to make connections between abstract mathematical concepts and practical situations [42]. Consequently, LLMs may broaden the scope of the subject and foster a deeper comprehension of mathematical principles.

Furthermore, there is a need to cultivate metacognition in mathematics education [46]. LLMs can contribute to enhancing metacognition, as they offer the capability to summarize knowledge related to a concept and provide examples that facilitate the transfer of this knowledge to different contexts. Students' metacognitive skills are further developed through their active inquiry and formulation of appropriate questions to engage with the large language model.

5 DISCUSSION

The rapid progress in AI and the emergence of LLM tools have attracted a lot of research interest in their potential as aids for teaching and learning. Recent studies demonstrate that LLMs, when utilized effectively, can provide valuable support to educators and students across various classes and subjects. While mathematics has received comparatively less attention thus far, it is by no means an exception to the potential benefits offered by these tools. The present study investigates the utilization of LLMs as alternatives or complements to the conventional tools that have traditionally been employed in mathematics education in recent decades. It argues that these innovative tools have the potential to provide functional and pedagogical opportunities that may influence changes in curriculum and assessment approaches.

Addressing a mathematical problem encompasses two essential dimensions: comprehending the underlying concepts and employing the appropriate mathematical calculations to arrive at solutions. While CASs have proven to be valuable tools for the latter, they often lack the ability to assist students in understanding the underlying concepts and determining the appropriate problem-solving steps.

This is where LLMs can complement CASs by enabling students to actively engage in dialogue with them, facilitating the process of solving mathematical problems. The explanations provided by an LLM regarding the underlying concepts, along with its guidance on problem-solving steps, have the potential to foster both students' conceptual understanding and their problem-solving skills. Nonetheless, as previously discussed, LLMs in their current state do not produce reliable calculation outcomes. This issue can be remedied by either manually validating the calculations with a CAS or by establishing an interface between the LLM and an online CAS. For instance, for ChatGPT, this can be accomplished using the plugin provided by Wolfram|Alpha. By harnessing these functional capabilities as well as their limitations, educators can reimagine and reshape both the content and assessment approaches to better align with the opportunities offered by LLMs, thereby enhancing the overall educational experience and outcomes for students.

The utilization of LLMs also holds significant pedagogical potential, impacting teaching and learning activities, approaches, and even the subject matter of Mathematics itself. In particular, LLMs facilitate activities that promote critical thinking, mathematical reasoning, and pen-and-paper skills in students. These activities involve observing, analyzing, and verifying the system's responses, as well as exploring the impact of parameters, discovering patterns, posing "what if…" questions, and identifying regularities. Furthermore, LLMs can be utilized in diverse ways for learning activities that leverage real-world contexts, serving to motivate and engage students in mathematical thinking, all while showcasing the practical applications and real-world uses of mathematics. Lastly, the versatility of LLMs in providing multiple representations can be leveraged in problem-solving activities, enabling students to interact with the system in textual, numerical, and symbolic forms, thereby enhancing their overall comprehension.

The integration of LLMs also allows for more student-centered, inquiry-based learning approaches. Educators may facilitate and guide students in using an LLM as a means to address research questions, test hypotheses, and discover knowledge. Differentiated learning is also automatically provided by LLMs since these systems may respond to each student or student group in an individualized way according to the dialogue established between them. In a similar manner, LLMs may also serve as scaffolding tools, offering each student tailored support to achieve their learning goals. This shift towards more constructivist approaches to teaching and learning also calls for changes in assessment that should focus more on the evaluation of higher-order thinking skills.

The subject matter of Mathematics, in terms of what is taught and in what depth, is also affected by the integration of LLMs in mathematics education. Educators may choose to further elaborate on mathematical concepts when addressing the calculation errors frequently made by LLMs in order to ensure students' understanding. LLMs also provide an opportunity for educators to move beyond the mere acquisition of procedural skills and rather emphasize the importance of deeply understanding mathematical concepts, their interrelationships, and their practical application in real-world contexts.

Despite the unique opportunities that LLMs offer, it is of critical importance to ensure their responsible and ethical use. This can be achieved by adhering to established guidelines and frameworks, such as the "Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators" [47] and "Artificial intelligence and future of teaching and learning: insights and recommendations" [48].

6 CONCLUSIONS

Over the past half-century, specialized software tools developed for mathematics have revolutionized teaching methodologies and redefined the roles of educators and students. The recent emergence of LLMs will likely have a similar, if not greater, impact. The previous sections discussed the functional and pedagogical opportunities that arise from harnessing LLMs in mathematics education, highlighting that their primary potential lies in offering students personalized learning experiences and real-time feedback while actively engaging them in dialogues regarding mathematical concepts and their practical application to real-world problems. These capabilities can serve as a means to motivate students, encourage their active participation, and cultivate their higher-order thinking skills, including problem-solving, critical thinking, mathematical reasoning, metacognition, and knowledge transfer across different contexts. On the other hand, issues and limitations currently associated with LLMs' use have also been identified. One such key issue is the computational errors in the results they produce, but this issue, if appropriately utilized by the educators, may prove to have a positive impact on students' learning.

Future research should focus on identifying the teaching and learning opportunities offered by LLMs, assessing their impact on learning outcomes, and developing effective strategies for their integration into curricula and course syllabi in higher education. The highly important concern of the responsible and ethical use of these innovative tools should also be investigated further, so as to ensure that they are employed in a considerate and safe way.

With research continuing [49], the increasingly widespread integration of LLMs in mathematics education, and the merging of mathematical software with AI systems, an unprecedented opportunity arises to enhance students' knowledge and skills in a world that is becoming increasingly digital and data-driven.

7 ACKNOWLEDGEMENT

The authors Nikolaos Matzakos and Maria Moundridou acknowledge financial support for the dissemination of this work from the Special Account for Research of ASPETE through the funding program "Strengthening ASPETE's research."

8 REFERENCES

[1] L. Etlinger, "The electronic calculator: A new trend in school mathematics," *Educational Technology*, vol. XIV, no. 12, pp. 43–45, 1974.

[2] R. Pierce and K. Stacey, "Mapping pedagogical opportunities provided by mathematics analysis software," *Int. J. Comput. Math Learning*, vol. 15, pp. 1–20, 2010. https://doi.org/10.1007/s10758-010-9158-6

[3] C. Ní Shé, E. Ní Fhloinn, and C. Mac an Bhaird, "Student engagement with technology-enhanced resources in mathematics in higher education: A review," *Mathematics*, vol. 11, no. 3, p. 787, 2023. https://doi.org/10.3390/math11030787

[4] T. Rasul *et al.,* "The role of ChatGPT in higher education: Benefits, challenges, and future research directions," *Journal of Applied Learning and Teaching*, vol. 6, no. 1, 2023. https://doi.org/10.37074/jalt.2023.6.1.29

[5] J. Moses, "Macsyma: A personal history," *Journal of Symbolic Computation*, vol. 47, no. 2, pp. 123–130, 2012. https://doi.org/10.1016/j.jsc.2010.08.018

[6] E. Roanes-Lozano, J. L. Galán-García, and C. Solano-Macías, "Some reflections about the success and impact of the computer algebra system DERIVE with a 10-Year time perspective," *Math. Comput. Sci*, vol. 13, pp. 417–431, 2019. https://doi.org/10.1007/s11786-019-00404-9

[7] A. J. Durán, M. Pérez, and J. L. Varona, "Misfortunes of a mathematicians' trio using computer algebra systems: Can we trust?" *Notices Amer. Math. Soc,* vol. 61, no. 10, pp. 1249–1252, 2014. https://doi.org/10.1090/noti1173

[8] A. Greiner-Petter *et al.,* "Comparative verification of the digital library of mathematical functions and computer algebra systems," in *Tools and Algorithms for the Construction and Analysis of Systems,* F. D. Rosu (Ed.). TACAS 2022, Springer, Cham, 2022, vol. 13243. https://doi.org/10.1007/978-3-030-99524-9_5

[9] R. Mayes, "Current state of research into CAS in mathematics education," in *The State of Computer Algebra in Mathematics Education*, J. Berry and J. Monaghan (Eds.). Chartwell-Bratt, 1997, pp. 171–189.

[10] J. B. Lagrange, "Using symbolic calculators to study mathematics: The case of tasks and techniques," in *The Didactical Challenge of Symbolic Calculators: Turning a Computational Device into a Mathematical Instrument*, 2005, pp. 113–135. https://doi.org/10.1007/0-387-23435-7_6

[11] Z. Lavicza, "The examination of Computer Algebra Systems (CAS) integration into university-level mathematics teaching," Unpublished Ph.D. dissertation. The University of Cambridge, Cambridge, UK.

[12] H. S. Risser, "What are we Afraid of? Arguments against teaching mathematics with technology in the professional publications of organisations for US mathematicians," *Int. J. Technol. Math. Educ*, vol. 18, no. 2, 2011.

[13] E. Akugizibwe and J. Y. Ahn, "Perspectives for effective integration of e-learning tools in university mathematics instruction for developing countries," *Education and Information Technologies*, vol. 25, no. 2, pp. 889–903, 2020. https://doi.org/10.1007/s10639-019-09995-z

[14] M. Thomsen, U. T. Jankvist, and K. M. Clark, "The interplay between history of mathematics and digital technologies: A review," *ZDM Mathematics Education,* vol. 54, pp. 1631–1642, 2022. https://doi.org/10.1007/s11858-022-01368-0

[15] M. Artigue, "Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work," *Int. J. Comput. Math. Learn,* vol. 7, pp. 245–274, 2002. https://doi.org/10.1023/A:1022103903080

[16] J. B. Lagrange and P. R. Richard, "Instrumental genesis in the theory of MWS: Insight from didactic research on digital artifacts," in *Mathematical Work in Educational Context: The Perspective of the Theory of Mathematical Working Spaces*, Cham: Springer International Publishing, 2022, pp. 211–228. https://doi.org/10.1007/978-3-030-90850-8_9

[17] J. B. Lagrange, "The present and future of AI in ME: Insight from empirical research," in *Mathematics Education in the Age of Artificial Intelligence. Mathematics Education in the Digital Era*, P. R. Richard, M. P. Vélez, and S. Van Vaerenbergh (Eds.). Springer, Cham, vol. 17, 2022, pp. 277–282.

[18] A. Ozkan, Z. Akcay, D. Boccio, D. Seo, and S. Budak, "Revising college algebra instruction to accept and incorporate the use of smartphone applications," in *ASEE Virtual Annual Conference Content Access*, 2021.

[19] S. Hadjinor, A. Asotigue, and A. Pangandamun, "Solving trigonometric problems using Mathway application in teaching mathematics," *Asian Journal of Research in Education and Social Sciences*, vol. 3, no. 3, pp. 87–97, 2021.

[20] J. Kenas, "Is using mathway cheating: All you need to know about mathway," 2021. [Online]. Available: https://t.ly/0TCA. [Accessed: June 23, 2023].

[21] N. B. B. Booc, K. Sobremisana, A. Ybañez, R. Tolosa, S. M. Ladroma, and K. M. Caparoso, "Artificial intelligence-powered calculator application usage in mathematics summative assessments," *Iconic Res. Eng. J*, vol. 6, no. 10, pp. 446–474, 2023.

[22] S. Wolfram, "About Wolfram|Alpha," Wolfram Alpha, 2020. [Online]. Available: https://www.wolframalpha.com/about/. [Accessed: June 23, 2023].

[23] S. Wolfram, "Wolfram|Alpha as the way to bring computational knowledge superpowers to ChatGPT," *Stephen Wolfram Writings*, 2023. [Online]. Available: https://writings.stephenwolfram.com/2023/01/wolframalpha-as-the-way-to-bring-computational-knowledge-superpowers-to-chatgpt/. [Accessed: June 23, 2023].

[24] V. E. Dimiceli, A. S. I. D. Lang, and L. Locke, "Teaching calculus with Wolfram|Alpha," *International Journal of Mathematical Education in Science and Technology*, vol. 41, no. 8, pp. 1061–1071, 2010. https://doi.org/10.1080/0020739X.2010.493241

[25] S. Abramovich, "On the interplay of mathematics and education: Advancing computational discovery from recognition to observation," *Mathematics*, vol. 10, no. 3, p. 359, 2022. https://doi.org/10.3390/math10030359

[26] M. Yildiz, "Computational thinking and geospatial intelligence in teacher education: Transdisciplinary approach," in *Proceedings of Society for Information Technology & Teacher Education International Conference*, E. Langran (Ed.). San Diego, CA, United States: Association for the Advancement of Computing in Education (AACE), 2022, pp. 117–125. [Online]. Available: https://www.learntechlib.org/primary/p/220717/. [Accessed: June 23, 2023].

[27] O. Knill, J. Carlsson, A. Chi, and M. Lezama, "An artificial intelligence experiment in college math education," [Online]. Available: http://www.math.harvard.edu/knill/preprints/sofia.pdf. [Accessed: June 23, 2023].

[28] B. A. Shawar and E. Atwell, "Chatbots: Are they really useful?" *Journal for Language Technology and Computational Linguistics*, vol. 22, no. 1, pp. 29–49, 2007. https://doi.org/10.21248/jlcl.22.2007.88

[29] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. North American Chapter Assoc. Comput. Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423

[30] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019. [Online]. Available: https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf. [Accessed: June 23, 2023].

[31] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint*, 2019. https://doi.org/10.48550/arXiv.1907.11692

[32] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *arXiv preprint*, 2019. https://doi.org/10.48550/arXiv.1910.10683

[33] A. Lewkowycz, A. Andreassen, D. Dohan, E. Dyer, H. Michalewski, V. Ramasesh, A. Slone, C. Anil, I. Schlag, T. Gutman-Solo, Y. Wu, B. Neyshabur, G. Gur-Ari, and V. Misra, "Solving quantitative reasoning problems with language models," *Advances in Neural Information Processing Systems*, 2022.

[34] S. Frieder, L. Pinchetti, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner, "Mathematical capabilities of ChatGPT," *arXiv preprint*, 2023. https://doi.org/10.48550/arXiv.2301.13867

[35] S. Imani, L. Du, and H. Shrivastava, "MathPrompter: Mathematical reasoning using large language models," *arXiv preprint*, 2023. https://doi.org/10.48550/arXiv.2303.05398; https://doi.org/10.18653/v1/2023.acl-industry.4

[36] Z. A. Pardos and S. Bhandari, "Learning gain differences between ChatGPT and human tutor generated algebra hints," *arXiv preprint*, 2023. https://doi.org/10.48550/arXiv.2302.06871

[37] S. Wolfram, "ChatGPT gets its 'Wolfram superpowers'!" *Stephen Wolfram Writings*, 2023. [Online]. Available: https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/. [Accessed: June 23, 2023].

[38] OpenAI, "ChatGPT plugins," 2023. [Online]. Available: https://openai.com/blog/chatgpt-plugins. [Accessed: June 23, 2023].

[39] S. Cameron, L. Ball, and V. Steinle, "The stability of mathematics students' beliefs about working with CAS," *Math Ed. Res. J*, 2023. https://doi.org/10.1007/s13394-023-00456-y

[40] D. Thurm and B. Barzel, "Teaching mathematics with technology: A multidimensional analysis of teacher beliefs," *Educ. Stud. Math*, vol. 109, pp. 41–63, 2022. https://doi.org/10.1007/s10649-021-10072-x

[41] A. Clark-Wilson, O. Robutti, and M. Thomas, "Teaching with digital technology," *ZDM Mathematics Education*, vol. 52, pp. 1223–1242, 2020. https://doi.org/10.1007/s11858-020-01196-0

[42] R. Pierce and K. Stacey, "Enhancing the image of mathematics by association with simple pleasures from real-world contexts," *Zentralblatt für Didaktik der Mathematik*, vol. 38, no. 2, pp. 214–225, 2006. https://doi.org/10.1007/BF02652806

[43] P. Hämäläinen, M. Tavast, and A. Kunnari, "Evaluating large language models in generating synthetic HCI research data: A case study," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, Hamburg, Germany, 2023, pp. 1–19. https://doi.org/10.1145/3544548.3580688

[44] R. Pierce, K. Stacey, R. Wander, and L. Ball, "The design of lessons using mathematics analysis software to support multiple representations in secondary school mathematics," *Technology*, *Pedagogy and Education*, vol. 20, no. 1, pp. 95–112, 2011. https://doi.org/10.1080/1475939X.2010.534869

[45] A. Dreher and S. Kuntze, "Teachers' professional knowledge and noticing: The case of multiple representations in the mathematics classroom," *Educ. Stud. Math*, vol. 88, pp. 89–114, 2015. https://doi.org/10.1007/s10649-014-9577-8

[46] A. Desoete and B. De Craene, "Metacognition and mathematics education: An overview," *ZDM Mathematics Education*, vol. 51, pp. 565–575, 2019. https://doi.org/10.1007/s11858-019-01060-w

[47] European Commission, Directorate-General for Education, Youth, Sport and Culture, "Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators," Publications Office of the European Union, 2022. [Online]. Available: https://data.europa.eu/doi/10.2766/153756. [Accessed: June 23, 2023].

[48] United States. Office of Educational Technology, "Artificial intelligence and the future of teaching and learning: Insights and recommendations report," Washington D.C., 2023. [Online]. Available: https://digital.library.unt.edu/ark:/67531/metadc2114121/m1/1/. [Accessed: June 23, 2023].

[49] R. Ghnemat, A. Shaout, and A. M. Al-Sowi, "Higher education transformation for artificial intelligence revolution: Transformation framework," *Int. J. Emerg. Technol. Learn*, vol. 17, no. 19, pp. 224–241, 2022. https://doi.org/10.3991/ijet.v17i19.33309

9 AUTHORS

Nikolaos Matzakos is an Assistant Professor of Applied Mathematics in the Department of Education at the School of Pedagogical and Technological Education (ASPETE) in Athens, Greece. His research interests include Nonlinear Differential

Equations and Inclusions, Computer Algebra Systems in Education and DL in Mathematics (E-mail: nikmatz@aspete.gr).

Spyridon Doukakis is a faculty member of the Bioinformatics and Human Electrophysiology Lab at the Department of Informatics at Ionian University, 7 Tsirigoti Square, 49100, in Corfu, Greece. His research interests include Computing Education, Educational Neuroscience, Distance Learning, and Knowledge Transformation (E-mail: sdoukakis@ionio.gr).

Maria Moundridou is an Assistant Professor of 'Educational Technology & Multimedia' in the Department of Education at the School of Pedagogical and Technological Education (ASPETE) in Athens, Greece. Her research interests include Intelligent Tutoring Systems, Adaptive Hypermedia, User Modelling, and Virtual Learning Environments. She has participated in a number of national and European R&D projects and published in prestigious journals and conference proceedings (E-mail: mariam@aspete.gr).