Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 18 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i18.51221

PAPER

Generative Artificial Intelligence on Mobile Devices in the University Preparation of Future Teachers of Mathematics

Peter Vankúš(✉)

Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava, Slovakia

peter.vankus@fmph.uniba.sk

ABSTRACT

The integration of artificial intelligence (AI) is revolutionizing the field of education. The possibilities of AI in education range from providing intelligent tutoring to facilitating computational thinking, a key skill in an AI-based society. As education systems around the world adapt to technological advances, the need for teachers to develop AI skills becomes crucial. This paper explores the use of AI on mobile devices to improve the preparation of future mathematics teachers. We acknowledge the need for teachers to be skilled in the use of AI tools to foster collaborative, effective, and ethical learning environments. We also provide a case study of our implementation of the generative AI ChatGPT in the preparation of future mathematics teachers at the Faculty of Mathematics, Physics and Computer Science at Comenius University in Bratislava. After this implementation, we analyzed students' opinions on the application of AI in selected areas. We present the results of this analysis and the implications for the practical use of AI in the university education of future mathematics teachers. Finally, based on our study results, we highlight the possibilities of the successful use of AI in different aspects of teaching and its temporary limitations, which are likely to be reduced by further technological developments.

KEYWORDS

generative artificial intelligence (AI), mathematics education, teacher preparation, mobile technology, educational research

1 INTRODUCTION

As defined by Balacheff [1], artificial intelligence (AI) refers to systems whose behavior appears intelligent to human observers. When examining an AI system, one can reasonably presume that its behavior results from some form of reasoning. A related theoretical goal is to model knowledge in a practical way. In the context of AI, intelligence essentially refers to the ability of implemented models

Vankúš, P. (2024). Generative Artificial Intelligence on Mobile Devices in the University Preparation of Future Teachers of Mathematics. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(18), pp. 19–33. https://doi.org/10.3991/ijim.v18i18.51221

Article submitted 2024-06-08. Revision uploaded 2024-07-28. Final acceptance 2024-07-29.

to solve problems. These solutions are not pre-encoded but are constructed by the machine itself.

According to Hwang and Tu [2], AI is the field of computer science devoted to creating systems capable of performing tasks that typically require human intelligence. These tasks include visual and speech recognition, reasoning, and decision-making.

Niemi [3] notes that while a universally agreed-upon definition of AI remains elusive, evidence from various sources highlights its significant impact on our lives. Despite differences in interpretation, common themes emerge: AI is computational intelligence. These intelligent machines analyze data, make inferences, and act autonomously [4]. Many definitions also characterize AI machines as learning entities capable of adapting to new tasks and making inferences through interactions with other data sources and humans.

1.1 Applications of artificial intelligence for education

The potential applications of AI in education are extensive. As outlined by Balacheff [1], the expectations for the effectiveness and efficiency of intelligent teaching and learning environments encompass several aspects. These include increasing the accessibility of knowledge, giving learners more autonomy, and potentially assisting or even replacing teachers in certain tasks. However, for teachers to successfully integrate new technologies into their daily practice, they need to be well informed about all aspects that determine their role within the didactic process. Understanding the computer-based system from a didactic perspective is essential, as is knowing a colleague with whom they might share responsibility for the class. Furthermore, the machine itself must be able to handle and produce relevant didactic information related to the teaching process. This interaction and collaboration between AI and teachers are an open challenge for both mathematics education studied and computer scientists. Meeting this challenge is crucial for the harmonious coexistence of AI and authentic teaching practices.

The UNESCO document [5] also highlights the challenge of preparing educators for an AI-driven education landscape while enabling AI systems to understand education. This dual process requires teachers to acquire new digital skills to effectively integrate AI into their teaching practices, while AI developers need to gain insight into educators' workflows and design solutions that remain viable in practical educational settings. This process of working with teachers to integrate AI into education is important. Teachers will continue to play a critical role in education, and the idea that AI can completely replace them is misguided. Opposing arguments oversimplify the teaching profession by focusing only on cognitive and repetitive tasks. They overlook studied that highlight the essential role of human mentors in supporting the learning process and neglect the creative and socio-emotional dimensions of teaching that go beyond the mere transmission of knowledge [6].

In addition, teachers should determine when and how to use AI-enabled tools. Therefore, the development and integration of these tools into educational programs should be a collaborative process that prioritizes the needs of educators over the assumptions of technologists or designers [7]. Nevertheless, AI technologies can automate routine administrative tasks such as grading and record-keeping, freeing up teachers' time to focus on the more inspirational and empathetic aspects of their profession.

1.2 Preparing teachers for the use of artificial intelligence in the classroom

To participate in the development of AI applications in education and to integrate such technologies into their teaching methods, teachers need to acquire specific AI-related skills. These include [7]:

- A clear understanding of how AI systems enhance learning, enabling informed judgments about new AI-based educational products.
- Study and data analysis skills to interpret insights from AI-generated data, ask relevant questions, and provide feedback to students.
- Enhanced management skills to oversee both human and AI resources.
- A critical perspective on the impact of AI and digital technologies on human life, fostering computational thinking and digital literacy among students.
- Leveraging AI to handle repetitive tasks allows teachers to focus on mentorship, emotional support, and interpersonal skills.
- Equipping learners with competencies that remain irreplaceable by machines.

As AI becomes more prevalent in the classroom, teacher training becomes a critical factor in empowering educators to use this technology. As Cukurova et al. [8] found in their sample of 800 teachers, an important factor in using AI in the classroom is ensuring that teachers are not burdened with additional workload. Equally important are teachers' understanding of AI, their confidence in using it, and their sense of ownership. Implementing support mechanisms for teachers when they need assistance and addressing ethical concerns are also essential. Building this understanding of AI and gaining the confidence to use it are also important goals for teachers' university preparation.

1.3 Study questions

Our study questions address the challenges of exploring the use of generative AI in the classroom, as outlined by Baidoo-Anu [9]:

- Using generative AI for student learning: How can we use generative AI to improve student learning? Should we train teachers and students to use existing generative AI tools to enhance teaching and learning?
- Integrating AI tools into teacher education programs: How can we integrate generative AI tools into teacher education programs? What strategies can prepare teacher candidates to effectively use AI tools in their classrooms?
- Addressing the digital divide [10]: Will these AI tools narrow or widen the existing digital divide? What steps can we take to move forward in an equitable way?

In our study, we will focus on the area of integrating AI tools into teacher education programs. At the Faculty of Mathematics, Physics, and Informatics of Comenius University in Bratislava, we decided to integrate generative AI into the preparation of future mathematics teachers. The students used this technology on the mobile devices available to them (smartphones, tablets, and notebooks). Our main goal was to answer the following study questions:

1. Will we be able to design and implement the use of generative AI technology in existing courses that will be beneficial to future mathematics teachers?
2. What will be the views of future mathematics teachers on the possibilities of the proposed implementations of AI in their future mathematics teaching?

By investigating these questions, we aim to support the integration of generative AI in the preparation of future mathematics teachers. The first question is related to testing specific applications of AI technology, and the second is related to building positive opinions of pre-service teachers about the possibilities of using it in their future teaching practice. The article is therefore addressed to professionals responsible for the preparation of future mathematics teachers and, by extension, to anyone interested in studying the use of generative AI in education.

2 METHODOLOGY

Our study has the character of a case study [11]. We integrated the generative AI ChatGPT into the preparation of future mathematics teachers at the Faculty of Mathematics, Physics, and Informatics of Comenius University in Bratislava. It was a first-year university course. The content of this course is probability and mathematical statistics. Students used AI technology through mobile devices such as smartphones, tablets, and laptops.

To answer the study questions, we studied the text outputs provided by the students who participated in the course in the years 2023 and 2024. A total of 40 students participated in the study (22 in 2023 and 18 in 2024). We have 11 text outputs produced by study participants in pair work from 2023 and seven outputs from 2024 (four students were absent during the class). In addition, we have 18 text outputs produced by students individually in 2024. Table 1 summarizes these numbers.

Table 1. Overview of analyzed students' text outputs

Year	Number of Participants	Text Outputs Produced in Pairs	Text Outputs Produced Individually
2023	22	11	–
2024	18	7	18

The individual assignment in 2024 required students to create four solved math tasks using ChatGPT on the following topics in probability and statistics: a) Basic concepts of probability and related problems (random phenomena, probability, complementary probability, independent phenomena, certain phenomena, impossible phenomena). b) Geometric probability and related problems. c) Conditional probability. d) Basic concepts of statistics: averages – arithmetic, weighted, geometric, harmonic, mode, median, and standard deviation. Students should thoroughly examine tasks designed by ChatGPT and judge the correctness of the given solution.

The assignment for the students in the pair work was formulated in three areas:

1. Solving the task. In 2023, students were given these instructions: "What is the probability of rolling a 6 at least once in 4 dice rolls?" In 2024, we provided less detailed instructions and allowed the students to investigate the mathematical task related to probability and statistics of their choice.
2. Demonstration and explanation of the concepts. "Ask ChatGPT to explain and demonstrate with an example the classical (Laplace's) definition of probability. Explain and demonstrate, with an example, Laplace's definition of probability at the elementary level. Explain and demonstrate with an example conditional probability." Again, in 2024, students were allowed to design their own inputs related to the explanation of concepts that they had deliberately chosen from the area of probability and statistics.

3. Test generation. "Prompt ChatGPT to create a test containing 10 multiple-choice questions about probability and statistics. At the end of the test, provide the correct answers and detailed solutions to each problem." This instruction was the same in 2023 and 2024.

The selection of these areas was based on the standard activities of mathematics teachers in the classroom, which are related to solving mathematical problems, explaining and demonstrating new concepts, and assessment, which is often related to test generation. In accordance with [8], we believe that the use of generative AI should be mainly to support existing activities in the classroom, not to invent new processes that could be perceived by teachers as time-consuming and demanding.

For each of these three areas, the following grading instruction was given: "Read the ChatGPT responses and grade them on a scale of A (excellent, outstanding results). B (very good, above average standard) C (good, normal, reliable work) D (satisfactory, acceptable results) E (adequate, results meet minimum criteria) FX (inadequate, further work required) based on the following criteria: a) correctness, b) completeness, c) appropriateness for the students." Therefore, the evaluation was in the form of a 6-point Likert scale. The names of the different scales corresponded to the students' course evaluations during their university studies. We chose such a scale because it is familiar to the students, which reduces the possibility of using the scale incorrectly.

Both the individual text documents and the pair of work text documents were subjected to content analysis [12] using the qualitative data analysis software ATLAS.ti. The results of the content analysis and our observations from the course are the basis for answering the first study question: "Will we be able to design and implement the use of generative AI technology in existing courses that will be beneficial to future mathematics teachers?"

The statistical analysis of the pair work text outputs serves to answer the second study question: "What will be the future mathematics teachers' views on the possibilities of the proposed implementations of AI in their future mathematics teaching?" Therefore, we quantitatively processed the data from the pair of work text documents and subjected them to statistical analysis. For this analysis, we used the nonparametric two-sided Fisher's exact test of independence, based on the ordinal categorical nature of our data and the need to compare their distribution with the assumed uniform distribution. The choice of this test was also influenced by expected values less than 5, which disqualified the use of the goodness-of-fit chi-square test or G-test of goodness-of-fit [13]. We will provide the results of the study in the next chapter of our paper.

3 RESULTS

3.1 Content analysis

In the content analysis, we worked with two different types of student-generated text documents. The first were documents created individually by 18 students in 2024. In these documents, the students provided four mathematical tasks on the topic of probability and statistics, which they created with the help of ChatGPT. The generative AI also provided the solutions. Students were asked to examine these tasks in detail and judge the correctness of the given solution. From the content analysis of the individually created text document in 2024, we obtained the codes described in Table 2.

Table 2. Codes in individual text document

Code	Number of Occurrences	Example
Correct solution	46	"ChatGPT designed and calculated this task correctly."
Minor mistake (correct procedure of solution, just small flaws)	5	"Correct solution, just could have made a more complete notation." "A bit unclear assignment (Why is the cube special? Are there 2 cubes or 1?)." "Procedure of solution good, but rounding is wrong, so the result is wrong."
Major mistake (bad solution or incorrect formulation of the task)	5	"The solution of the mean, modus, and median is correct, but ChatGPT only put the variance calculation into an incompletely written fraction and did not complete its calculation, and it only explained and did not calculate the standard deviation." "ChatGPT could not find me an example involving geometric probability despite several attempts, it always based the assignment on geometric shapes, geometric sequence but the example was calculated by basic probability." "The assignment is good, but the solution is not."

From the codes, we can see that the prevailing students' evaluation of the ability of the generative AI ChatGPT to correctly design and solve the mathematical tasks in probability and statistics was positive. Some students' comments showed that they were even surprised by the ability of generative AI in this role. "ChatGPT did the job very well, I'm surprised :D," "ChatGPT was a pleasant surprise in this case and will surely come in handy for similar work in the future."

The minor mistakes that students observed could have been potentially improved in the next conversations with the AI and through iterative improvements to its answers. The major mistakes could be partially due to the technical problems of the ChatGPT in solving mathematical tasks [14]. This situation could be improved by the progress of generative AI's abilities to solve mathematical problems in the future [15].

As for the content analysis of the pair of work text documents in 2023, we found two positive reactions of students to the ability of AI to solve mathematical tasks in probability and statistics. Students stated: "ChatGPT calculated the task well and explained it clearly," and "Both the procedure and the result are suitable for the pupils, the procedure is well and clearly written (although the mathematical language is sometimes clumsy)."

As for the explanation and demonstration of the concepts in probability and statistics, all two students' reactions were also positive: "Well explained, clear complete," and "ChatGPT explained it correctly and also gave a good example."

The ability to generate the test from the probability and statistics topics using ChatGPT was rated poorly in all three comments: "Tasks are good for inspiration, but the results may not be correct." "ChatGPT is not able to create a suitable test with correct answers and solutions; for some tasks, it wrote the correct solution procedure, but in the correct answers, something else was marked as the correct answer; it could only serve as inspiration when creating a test, but it is necessary to check the correctness of the tasks." and "This test might be usable after careful checking and correction, but in its 'raw' form (as Chat created it), it is not very usable." In the content analysis of the pair work documents from 2024, we found only three comments connected with the test generation: "The tasks were often similar," "Tasks do not make sense or are incomplete; a lot of information is missing." and "Some problems were poorly formulated; others had good solutions but the wrong answer."

The content analysis of the pair work text documents from the point of view of the students' opinions about the possibilities of using generative AI in teaching probability and statistics was limited by the fact that only a few documents have

some additional comments besides the scale ratings. Therefore, to answer the second study question, the following statistical analysis of the scale ratings will be more informative.

3.2 Statistical analysis

Working in pairs, the math education students used the generative AI ChatGPT to produce outputs in text document format in three areas: 1) solving the task, 2) demonstration and explanation of the concepts, and 3) test generation. These areas are described in detail in the methodology section of the paper. In each of these areas, the students evaluated the ChatGPT output according to three criteria: a) correctness, b) completeness, and c) appropriateness for the students. In the following part of the paper, we provide a sample student text for each of these areas, and then summarize the data collected from the students.

Solving the task. The assignment for the students in this area was in the year 2023: "Using ChatGPT, try to answer: What is the probability of rolling a six at least once in four dice rolls?" In the year 2024, students could use ChatGPT to find the answers to their own mathematics tasks in probability and statistics. An example of the students' solution is depicted in Figure 1.

What is the probability of rolling a 6 at least once in 4 dice rolls?

On one roll of the dice, the probability of rolling a 6 is 1/6. Therefore, the probability of not rolling a 6 in one roll of the dice is equal to 5/6.

Since the dice rolls are independent, the probability that all four dice rolls do not roll a 6 is the product of the probability that a 6 does not fall in each of the four dice rolls:

(5/6) * (5/6) * (5/6) * (5/6) = 625/1296

However, since we are interested in the probability that a 6 will fall at least once, we can calculate the inverse probability - that is, the probability that none of the rolls will be a 6, and then subtract it from 1 to get the desired probability:

1 - 625/1296 = 671/1296

So the probability of rolling a 6 at least once in four rolls of the dice is approximately 0.517, or 51.7%.

(a) correctness - B

(b) completeness - B

(c) suitability for pupils – C

Fig. 1. Part of students' solution, answering the questions with ChatGPT

Results of students' feedback in Area 1. Solving the tasks is shown in Table 3 (year 2023) and Table 4 (year 2024). The table shows the number of students selecting the given scale. A (excellent, outstanding results) B (very good, above average standard) C (good, normal, reliable work) D (satisfactory, acceptable results) E (adequate, results meet minimum criteria) FX (inadequate; further work is required).

The table is divided by criteria, such as (a) correctness, (b) completeness, and (c) appropriateness for the pupils. It also includes the expected values for a uniform distribution of responses and a statistical evaluation of the differences between the actual and expected numbers based on a nonparametric Fisher's exact test of independence.

Table 3. Results in area solving the tasks, 2023

1. Correctness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	9	2	0	0	0	0
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	**p-value = 0.0093**					
2. Completeness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	9	2	0	0	0	0
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	**p-value = 0.0093**					
3. Appropriateness for the pupils						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	7	3	1	0	0	0
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	p-value = 0.0843					

Table 4. Results in area solving the tasks, 2024

1. Correctness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	5	0	0	0	1	1
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.2890					
2. Completeness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	5	0	0	1	0	1
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.2890					
3. Appropriateness for the pupils						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	5	0	0	0	1	1
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.2890					

Demonstration and explanation of the concepts. In 2023, the assignment for students in this area was: "Using ChatGPT, explain and demonstrate with an example the classical (Laplace's) definition of probability. Explain and demonstrate, with an example, Laplace's definition of probability at the elementary school level. Explain and demonstrate with an example conditional probability." In 2024, students

were able to choose their topics from the areas of probability and statistics. An example of one part of the students' solution is in Figure 2.

Explain and demonstrate with an example Laplace's diagram at the level of an elementary school student. Grade: A, A, A.

Laplace's diagram can also be used to help us understand probability. Probability is a way of measuring how likely something is to happen. For example, we can use probability to see how likely it is that a certain number will roll on a die.

Laplace's scheme in probability is to count the number of possible outcomes and divide that number by the total number of outcomes. For example, if we roll an ordinary dice with six sides, there are four possible outcomes (rolls of 1, 2, 3, 4, 5 or 6). The probability that any of these numbers will be rolled is 1/6, because there is only one way to roll a particular number out of the six possible outcomes.

Fig. 2. Part of students' solution, explanation of concepts with ChatGPT

Results of student feedback in Area 2. *Demonstration and explanation of the concepts* are presented in Table 5 (for the year 2023) and Table 6 (for the year 2024). These tables display the number of students selecting each grade on the given scale: A (excellent, outstanding results), B (very good, above average standard), C (good, normal reliable work), D (satisfactory, acceptable results), E (adequate, results meet minimum criteria), and FX (inadequate, further work required). The data is categorized by criteria: (a) correctness, (b) completeness, and (c) appropriateness for the pupils. Additionally, the tables include the expected values assuming a uniform distribution of responses and a statistical evaluation of the differences between the actual and expected numbers using a nonparametric Fisher's exact test of independence.

Table 5. Results in area demonstration and explanation of the concepts, 2023

1. Correctness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	8	1	2	0	0	0
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	**p-value = 0.0496**					
2. Completeness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	5	5	1	0	0	0
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	p-value = 0.1038					
3. Appropriateness for the pupils						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	3	4	3	1	0	0
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	p-value = 0.5539					

Table 6. Results in area demonstration and explanation of the concepts, 2024

1. Correctness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	4	1	1	1	0	0
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.8601					
2. Completeness						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	3	2	2	0	0	0
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.6853					
3. Appropriateness for the pupils						
Responses	**A**	**B**	**C**	**D**	**E**	**FX**
Observed frequency	4	1	1	1	0	0
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.8601					

Test generation. The assignment for the Area 3. *Test Generation* was in both years 2023 and 2024: "Using ChatGPT, create a test containing 10 multiple-choice problems on probability and statistics. At the end of the test, provide the correct answers and detailed solutions to each problem."

As in the previous areas, the students graded the outputs on a scale of A (excellent, outstanding results), B (very good, above average standard). C (good, normal, reliable work) D (satisfactory, acceptable results) E (adequate, results meet minimum criteria) FX (inadequate, further work required) from the point of a) correctness, b) completeness, and c) appropriateness for the pupils.

The example of one part of the students' solution is depicted in Figure 3, and the results together with their statistical analysis are in Table 7 for the year 2023 and in Table 8 for the year 2024.

Create a test containing 10 multiple-choice problems on probability and statistics. At the end of the test, give the correct answers and detailed solutions to each problem.

1) What is the probability that the number 3 or 5 will be rolled when the dice are rolled?

a) 1/3 b) 1/6 c) 1/2 d) 1/4

Answer: b) 1/6

Solution: A die has 6 sides; each side has one number. The probability of hitting 3 or 5 is 2/6, which can be simplified to 1/3.

Fig. 3. Part of students' solution, test generation

Table 7. Results in area test generation, 2023

1. Correctness						
Responses	A	B	C	D	E	FX
Observed frequency	1	1	4	2	0	3
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	p-value = 0.7657					
2. Completeness						
Responses	A	B	C	D	E	FX
Observed frequency	2	3	1	1	0	4
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	p-value = 0.7657					
3. Appropriateness for the pupils						
Responses	A	B	C	D	E	FX
Observed frequency	1	1	3	1	0	5
Assumed distribution	1.83	1.83	1.83	1.83	1.83	1.83
Fisher's test of independence	p-value = 0.6086					

Table 8. Results in area test generation, 2024

1. Correctness						
Responses	A	B	C	D	E	FX
Observed frequency	2	0	1	2	1	1
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 1.0000					
2. Completeness						
Responses	A	B	C	D	E	FX
Observed frequency	1	3	1	0	1	1
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.9441					
3. Appropriateness for the pupils						
Responses	A	B	C	D	E	FX
Observed frequency	1	3	1	1	0	1
Assumed distribution	1.17	1.17	1.17	1.17	1.17	1.17
Fisher's test of independence	p-value = 0.9441					

4 DISCUSSION

4.1 Discussion of study results

The discussion will be divided into two parts based on our study questions:

1. Will we be able to design and implement the use of ChatGPT generative AI technology in existing courses that will be beneficial to future mathematics teachers?

2. What will be the future mathematics teachers' views on the possibilities of the proposed implementations of AI in their future mathematics teaching?

Considering the first study question, we successfully implemented generative AI in our course in probability and statistics. During the experimental activities, students used the AI in three main areas: 1. solving (and generating) the tasks; 2. demonstration and explanation of the concepts; and 3. test generation. We selected these areas as they relate to the common activities that our students, as future mathematics teachers, would use in their teaching. Based on the evidence from the content analysis, the students found our demonstration of the capabilities of generative AI beneficial for their future practice. E.g., they wrote: "ChatGPT came as a pleasant surprise and will certainly be useful in the future when designing similar tasks."

According to the second study question, "What will be the future mathematics teachers' views on the possibilities of the proposed implementations of AI in their future mathematics teaching?" From the result obtained by the analysis of pair work text documents, we can see that students see positive implementations of AI in the area of solving tasks. Students stated: "ChatGPT calculated the task well and explained it clearly," and "Both the procedure and the result are suitable for the pupils; the procedure is written well and clearly (although the mathematical language is sometimes clumsy)." In all categories, the modus values had the highest scale rating of "A" (2023 and 2024). Students' opinions were statistically significantly different ($p < 0.05$) from the uniform distribution in the categories "completeness" and "correctness" in 2023; the other differences are not statistically significant.

Also, in the area of demonstration and explanation of the concepts, all students' comments were positive: "well explained, clear, and complete," "ChatGPT explained it correctly and also gave a good example." The modus values of scale ratings were in this area the highest scale rating "A," but the differences between the students' data and the theoretical uniform distribution were not statistically significant in any category both in 2023 and 2024.

In the area of test generation, all students' comments indicated poor performance, e.g., "Some problems were poorly formulated; others had good solutions but the wrong answer." The modus scale rating in this area in the year 2023 was the worst "FX" in all categories. In the year 2024, the modus values were "A" and "D" in the category of "correctness" and "B" in both categories "completeness" and "appropriateness for the pupils." Again, these differences are not statistically significant.

Also, from the content analysis of the individual students' text outputs from 2024, we can see that students highly appreciate the ability of generative AI to assist them in generating and solving the mathematical tasks from the selected themes of probability and statistics. For example, they stated, "The tasks from ChatGPT were surprisingly appropriate, clear, and correct. All the tasks were solved right away. Here and there, we found minor errors that needed to be corrected, but very rarely. ChatGPT managed to solve all the tasks correctly."

4.2 Limitations and implications for future study

A limitation of our study is the limited size of the study sample, which consisted of students of our probability and statistics course for first-year preservice mathematics teachers at the Faculty of Mathematics, Physics, and Informatics of Comenius University in Bratislava. Another specificity is the form of implementation of the use of generative AI ChatGPT in the classroom, which influenced the

resulting students' opinions. Therefore, our conclusions cannot be applied to the general population. They mainly serve as an example of good practice in the use of generative AI in the preparation of future mathematics teachers and point to areas where, according to our results, it is possible to use these technologies successfully or to areas where the results are so far worse. To draw more general conclusions, the study needs to be conducted with a larger sample of students and with more types of activities in which generative AI is implemented. Such a future study will provide a broader view of the potential of generative AI in the preparation of future mathematics teachers.

5 CONCLUSION

As a UNESCO document [5] states, education has a critical role to play in preparing the future workforce for generative AI applications. Closing the AI skills gap requires more than the adoption of powerful learning technologies. It requires a re-evaluation of teaching content and methods at all levels of education. There is a need for AI competencies that go beyond basic IT skills. Rather than limiting these competencies to 21st-century skills, countries should focus on equipping learners with the ability to identify and solve problems using computational techniques and technologies.

This goes hand in hand with the preparation of future teachers. In line with Lee [16], we agree that one of the challenges in integrating AI into education is the professional development of teachers in this area. However, the field is still in the early stages of understanding the necessary knowledge, skills, and pedagogical practices that teachers need to implement generative AI effectively [17]. In this context, our paper presents the successful integration of the generative AI ChatGPT used on mobile devices in the training of future mathematics teachers. Students used the AI in three main areas, and we analyzed students' assessments of the AI's performance. These areas were focused on the jobs that are highly relevant for every future mathematics teacher: solving and generating mathematical tasks, demonstrating and explaining mathematical concepts, and generating and solving tests to evaluate students' mathematics knowledge.

The choice of these areas was deliberately made with the intention that they should be existing components of school education, not redundant activities just for the purpose of integrating AI. In our opinion, which corresponds with the experts [5] and [8], to successfully integrate generative AI into the classroom, it is necessary to ensure that its applications really facilitate the normal work of teachers and do not add unnecessary extra burden.

Our results show that students perceived the abilities of generative AI positively for helping them solve mathematical tasks in the areas of probability and statistics. They highly value the correctness and completeness of AI solutions. Also, the most prevalent are positive students' views on the ability of AI to demonstrate and explain mathematical concepts. In the area of test generation, most students deemed the capability of AI to be very bad in 2023, but the results were better in 2024. This could be thanks to the increasing capabilities of ChatGPT in undergraduate mathematics [18]. The results concerning the AI's abilities in solving the tasks are statistically significant; in the next two areas, the differences are not statistically significant. This relates to the limitations of our study, the limited size of the study sample, and specific implementations of AI. Thus, there is a need for future studies with a larger study sample and incorporating more activities with the use of

generative AI. Interesting would also be a long-term study concerning how our students will implement AI in their future practices.

6 ACKNOWLEDGMENTS

The study was financially supported by The Ministry of Education, Research, Development and Youth of the Slovak Republic; grant number KEGA 037UK-4/2024: Innovative learning technologies in the preparation of future mathematics teachers.

7 REFERENCES

[1] N. Balacheff, "Artificial intelligence and mathematics education: Expectations and questions," in *Proc. 14th Bienn. Conf. Australian Assoc. Math. Teachers*, Perth, Australia, 1993, pp. 1–24.

[2] G.-J. Hwang and Y.-F. Tu, "Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review," *Mathematics*, vol. 9, no. 6, p. 584, 2021. https://doi.org/10.3390/math9060584

[3] H. Niemi, "AI in learning: Preparing grounds for future learning," *Journal of Pacific Rim Psychology*, vol. 15, 2021. https://doi.org/10.1177/18344909211038105

[4] J. Roschelle, J. Lester, and J. Fusco, Eds., "AI and the future of learning: Expert panel report," *Digital Promise*, 2020. https://circls.org/reports/ai-report

[5] "Artificial intelligence in education: Challenges and opportunities for sustainable development," UNESCO, 2019. https://www.gcedclearinghouse.org/sites/default/files/resources/190175eng.pdf

[6] M. Bali, "Against the 3A's of EdTech: AI, Analytics, and Adaptive Technologies in Education," The Chronicle of Higher Education, 2017. https://www.chronicle.com/blogs/profhacker/against-the-3as-of-edtech-ai-analytics-and-adaptive-technologies-in-education/64604

[7] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, *Intelligence Unleashed. An Argument for AI in Education*. London: Pearson, 2016.

[8] M. Cukurova, "The interplay of learning, analytics, and artificial intelligence in education," 2024. [Online]. Available: https://arxiv.org/abs/2403.16081

[9] D. Baidoo-Anu and L. Owusu Ansah, "Education in the era of generative Artificial Intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," *Journal of AI*, vol. 7, no. 1, pp. 52–62, 2023. https://doi.org/10.61969/jai.1337500

[10] J. A. G. M. Van Dijk, "Digital divide research, achievements and shortcomings," *Poetics*, vol. 34, nos. 4–5, pp. 221–235, 2006. https://doi.org/10.1016/j.poetic.2006.05.004

[11] R. K. Yin, *Case Study Research and Applications: Design and Methods*, 6th ed. Thousand Oaks, CA: Sage, 2018.

[12] M. Schreier, "Content analysis, qualitative," in *SAGE Research Methods Foundations*, P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, and R. A. Williams, Eds., 2019. [Online]. https://doi.org/10.4135/9781526421036753373

[13] J. H. McDonald, *Handbook of Biological Statistics*, 3rd ed. Baltimore, MD: Sparky House Publishing, 2014.

[14] S. Schorcht, N. Buchholtz, and L. Baumanns, "Prompt the problem – investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques," *Front. Educ.*, vol. 9, 2024. https://doi.org/10.3389/feduc.2024.1386075

[15] K. Faldu, A. Sheth, P. Kikani, M. Gaur, and A. Avasthi, "Towards tractable mathematical reasoning: Challenges, strategies, and opportunities for solving math word problems," *arXiv preprint arXiv:2111.05364*, 2021. https://arxiv.org/abs/2111.05364

[16] I. Lee and B. Perret, "Preparing high school teachers to integrate AI methods into STEM classrooms," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, no. 11, pp. 12783–12791. https://doi.org/10.1609/aaai.v36i11.21557

[17] S. Kim *et al.*, "Analyzing teacher competency with TPACK for K-12 AI education," *KI – Künstliche Intelligenz*, vol. 35, no. 2, pp. 139–151, 2021. https://doi.org/10.1007/s13218-021-00731-9

[18] S. Frieder *et al.*, "Mathematical capabilities of ChatGPT," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36, 2023, pp. 27699–27744.

8 AUTHOR

Peter Vankúš is a Lecturer at the Faculty of Mathematics, Physics, and Informatics at Comenius University in Bratislava, Slovakia. His research focuses on innovative technologies in mathematics education. In particular, he promotes the use of educational games and artificial intelligence in the classroom and in the preparation of future mathematics teachers (E-mail: peter.vankus@fmph.uniba.sk).