Mathematics and Critical Thinking in the AI ERA: Rethinking Classroom Practices

Cristina P. S. Dias
cpsd@ipportalegre.pt
Portalegre Polytechnic University
Portalegre, Portugal
NOVAMATH – Center for Mathematics and Applications
New University of Lisbon, Portugal

Luísa M. S. Carvalho
luisacarvalho@ipportalegre.pt
Center for Research in Education and Psychology of the University of Évora (CIEP-UE)
Évora, Portugal
CARE – Research Center on Health and Social Sciences
Portalegre, Portugal

Sérgio D. Correia
scorreia@ipportalegre.pt
Portalegre Polytechnic University
Portalegre, Portugal
CARE – Research Center on Health and Social Sciences
Portalegre, Portugal

Abstract

In an educational scenario marked by the growing presence of Artificial Intelligence (AI), mathematics plays a strategic role in shaping critical thinking and solving complex problems, ensuring an inclusive approach. Overcoming traditional approaches centered on formulas and mechanical procedures, while promoting inclusiveness, is a pedagogical priority. This study aimed to implement and analyze two didactic proposals that integrate AI tools into problem-solving situations in 120-minute classes designed for higher education students. The methodology adopted involved, in the first session, guided exploration of an AI tool, with a focus on formulating questions, analyzing answers, and the strategic use of keywords. In the second session, the students applied mental calculation strategies supported by AI, solved problems in groups, and were assessed based on the application of strategy, interpretation, and collaboration. The results indicate a strengthening of students' autonomy, an improvement in problem-solving skills, and greater critical engagement in the use of AI to support mathematical reasoning.

Keywords

Education, Artificial Intelligence in Learning, Problem-Solving Strategies, Critical Thinking Development

1 Introduction

The COVID-19 pandemic has accelerated the transition to digital teaching and learning environments, driving the integration of emerging technologies into teaching practices. Among these, artificial intelligence (AI) has emerged as a promising tool for personalizing learning, supporting formative assessment, and promoting student autonomy. Integrating AI-based educational tools not only holds advantages for students, educators, and institutions, but also has the potential to facilitate or notably enhance inclusive education [25]. Digital technologies have revolutionized the landscape of Higher Education Institutions (HEIs), enhancing accessibility and facilitating innovative teaching and learning methods [7], however, many teachers have been faced with the challenge of using these technologies with little or no prior training [5, 17]. Although platforms such as Coursera, edX, and Khan Academy have begun to offer training on AI in education, a significant gap remains in defining competency benchmarks for teachers in this area. The research proposed in [9] underscores the need to develop competency frameworks that encompass not only the technical mastery of AI but also its pedagogical applications and the ethical implications of its use in a school context. In the teaching of mathematics, the use of AI represents a strategic opportunity to transform the learning experience [6]. Tools such as ChatGPT and Microsoft Copilot can serve as virtual tutors, providing step-by-step explanations, answering questions, and promoting accessible and personalized language [15]. This technological mediation enables students to progress at their own pace, cultivate greater confidence, and reinforce their understanding of concepts that are often considered abstract or challenging [4]. Additionally, AI can facilitate the development of higher-order skills, including critical thinking, creativity, and collaborative problem-solving. As *Skovsmose* argues [23], 21st-century mathematics education must go beyond the mechanical execution of procedures, emphasizing interpretation, argumentation, and informed decision-making. In this context, it is essential to explore the potential of AI tools as catalysts for more critical and innovative educational practices. This study aims to analyze the integration of ChatGPT and Microsoft Copilot in statistical problem-solving situations, with a focus on promoting critical thinking, autonomy, and the ability to formulate and test statistical hypotheses. The research was developed around three problem situations worked on with higher education students.

2 Artificial Intelligence and Critical Thinking: Risks, Challenges, and Educational Implications

AI has emerged as a powerful and transformative tool, with the potential to amplify human cognition, personalize learning, and support decision-making processes [19, 21]. Integrating AI-based educational tools not only offers advantages for students, educators, and institutions, but also has the potential to facilitate or enhance inclusive education, as noted in [26]. However, the growing ubiquity of AI in everyday life raises pressing challenges, particularly regarding its impact on our ability to employ critical thinking. This uniquely human skill involves interpreting, evaluating, and questioning information in a reasoned and reflective manner [11]. In a world where algorithms suggest what to eat, where to go, or what to believe, there is a risk of delegating essential cognitive processes to automated systems. Over-reliance on AI tools can lead to a decline in students' critical thinking skills, as they may begin to accept AI-generated answers without critically evaluating them [1, 24]. Blind trust in AI suggestions,

Information Society 2025, Ljubljana, Slovenia

such as virtual assistants or predictive applications, can compromise our cognitive autonomy and foster mental passivity [22]. For example, by accepting suggestions without checking them, we stop considering alternatives or constructing our reasoning. This attitude, although comfortable, weakens the skills that support conscious and responsible decisions. Even more worrying is the potential for AI to distort our perception of reality. The emergence of deepfakes, artificially generated texts, images, or videos with convincing realism, raises serious concerns about our ability to distinguish the real from the manipulated [12]. These technologies can contribute to the spread of disinformation and also undermine social and interpersonal trust. Furthermore, AI can reduce students' ability to think innovatively and creatively, replacing original thinking with the standardised reasoning of AI [1, 10]. For all these reasons, there is an urgent need to educate people to use AI critically, consciously, and ethically. Training in AI literacy is essential to enable the informed and critical use of AI tools [13, 29], considering AI as a support tool rather than a substitute for human thought. It is essential to develop the ability to question the answers that AI provides, to draw on diverse sources, and to reflect on the origins, limits, and biases of algorithmic systems. This practice is indispensable for preserving mental agility and self-reliance, particularly among younger generations. In the educational context, particularly in teaching institutions, it is essential to promote the development of critical thinking skills so that students can distinguish between true and false information, especially in an environment where AI is widely used [20]. AI should be used to stimulate reasoning, not to replace it. Tools such as ChatGPT or Microsoft Copilot can act as cognitive mediators when integrated pedagogically. They allow hypotheses to be tested, strategies to be compared, and feedback to be given. As Trikoili et al. [27] argue, combining human and AI assessments can be a practical approach to ensuring that critical thinking skills are accurately assessed, leveraging the strengths of both methods. Students should be encouraged to critically evaluate their answers, justify their decisions, and discuss different approaches to solving problems. These practices align with the concept of critical mathematical literacy [23, 14], which extends beyond technical fluency to encompass the values of interpretation, argumentation, and decision-making.

It's also important to remember that AI is developed, trained, and tested by humans. Scientists and engineers define the data to be used, build the algorithms, analyze the results, and determine their applications. Human subjectivity is present at every stage of the process, influenced by cultural, ethical, and epistemological contexts [18]. This means that, even in a world increasingly shaped by machines, human responsibility remains non-negotiable. Preparing students to act in this scenario requires cultivating skills that enable them to interact with intelligent systems in an ethical and critical manner [19]. Schools must promote the ability to reflect on the role of technology in shaping knowledge, relationships, and the collective future. The aim is not to reject AI, but to integrate it consciously so that it becomes a collaborator, not a substitute, in the development of autonomous, creative, and thinking citizens.

3 Methodology

This research is part of a qualitative paradigm, with a descriptive and exploratory nature, aiming to understand the dynamics of teaching and learning in pedagogical practices supported by AI tools within the context of mathematics education. As Bogdan and Biklen [3] point out, qualitative research focuses on analyzing the meanings attributed by participants to their experiences through direct observation and detailed description of educational processes in natural contexts. The study was conducted over two 120-minute teaching sessions, implemented in a higher education class comprising approximately 15 students. The intervention plan was designed to integrate AI tools, specifically ChatGPT and Microsoft Copilot, into statistical problem-solving activities, promoting autonomy, critical thinking, and collaborative reflection. In the first session, students were instructed to interact with the AI to formulate transparent and objective questions, identify relevant keywords, and critically interpret the answers generated. The second session focused on the application of mental calculation strategies, with the support of the AI, allowing the comparison of different resolution methods in group work. Data collection was based on direct observation, supported by a structured grid, the teacher's field notes, and student output, including written responses, interactions with the AI, and justifications for the strategies employed. For the formative assessment, an analytical rubric specifically designed for this practice was used to evaluate the application of mathematical strategies, the interpretation of results, and peer collaboration. The data was analyzed using an interpretive approach [8] to identify patterns of mathematical thinking, levels of autonomy, attitudes towards the use of AI, and signs of critical thinking in statistical decision-making. This methodology enabled the understanding not only of the products of learning but also of the cognitive and social processes involved in integrating AI into teaching contexts.

3.1 Description of Pedagogical Practice

This proposal was designed for two 120-minute sessions, held in a room equipped with computers and internet access. The primary objective of the practical was to promote an understanding and application of the concepts associated with hypothesis testing through an interactive, contextualized approach mediated by artificial intelligence tools. The first lesson focused on introducing the fundamental concepts of statistical inference. The students were organized into small groups and initially took part in an introductory activity to explore their preconceptions about hypothesis testing. This was followed by a discussion led by the teacher, who introduced and clarified the main theoretical elements: the null hypothesis () and the alternative hypothesis (H_1), type I and type II errors, the significance level, the test statistic, and the p-value. This stage was supported by the ChatGPT and Microsoft Copilot tools, used for the following purposes: (a) to reformulate and clarify concepts present in the textbooks, (b) to generate explanatory examples and even intentionally incorrect examples to promote discussion, (c) to illustrate how AI can contribute to the interpretation of statistical decisions. Students were challenged to ask AI questions, such as: (1) How do I know if I reject the null hypothesis? (2) What does a p-value of 0.03 mean? (3) Can the sample mean be used to reject H_0 in this case?

This first session aimed to familiarize students with the statistical language of hypothesis testing while promoting a critical and reflective use of the answers generated by AI, discussing their relevance, correctness, and conceptual clarity. The second lesson consisted of the practical application of the knowledge acquired, with a focus on solving statistical problem situations, using collaborative work. Each group received a set of simulated (or real) data and a statement with an inferential question, such as:

- *Do school students sleep less than 7 hours a night on average?*
- *Is there evidence that the proportion of users satisfied with AI exceeds 60%?*

Based on the data provided, the groups had to: (i) formulate the null hypothesis (H_0) and the alternative hypothesis (H_1), (ii) define the appropriate significance level (e.g. 0.05), (iii) calculate or interpret the test statistic and p-value, (iv) make a reasoned inferential decision, (v) use ChatGPT or Copilot to confirm the reasoning and explore alternative explanations or validate the conclusions generated with AI support.

Throughout the process, the teacher took on the role of facilitator, circulating among the groups, promoting debate, answering conceptual questions, and encouraging comparison between their own resolution and the resolution proposed by the AI.

3.2 Practice evaluation

The assessment was formative, guided by explicit criteria, which valued: (a) the correct formulation of hypotheses, (b) the appropriate interpretation of the p-value, (c) the apparent justification of the statistical decision made, (d) the critical and conscious use of AI tools (avoiding automatic or uncritical responses), (e) the clarity and rigor in communicating the results.

To organize the teaching practice, the teacher distributes three problem situations to the groups to work on the topic of Hypothesis Testing. The three problem situations include simulated data and clear questions, ready to be used either on paper or using ChatGPT and Copilot. The first problem situation deals with students' hours of sleep:

- *The aim is to find out whether students at a school sleep, on average, less than 7 hours a night. A random sample of 20 students was taken.*

Data to be used in the first problem situation (hours of sleep per night): 6.2, 5.8, 7.1, 6.5, 6.0, 6.4, 5.7, 6.8, 5.9, 7.0, 6.3, 6.6, 5.5, 6.1, 6.9, 5.8, 6.7, 6.0, 6.2, 5.6

- *Based on this data, is there statistical evidence that students sleep less than 7 hours a night?*

Instructions given by the teacher: (1) formulate the hypotheses H_0 and H_1; (2) consider the significance level $\alpha = 0.05$; (3) use AI to calculate the p-value or compare it with the critical value; (4) decide whether or not to reject H_0; (5) justify your decision with the support of AI (you can ask ChatGPT or Copilot to run the t-test for this sample and interpret the result).

For the second Problem-situation (satisfaction with an AI app), the following statement is given:

- *"A company wants to know if more than 60% of users are satisfied with its new AI application."*

Consider a sample of 150 users, where 102 reported being satisfied. Does the data suggest that the proportion of satisfied users is higher than 60%?

Instructions given by the teacher: (1) formulate the hypotheses H_0 and H_1; (2) consider the significance level $\alpha = 0.05$; (3) ask the AI to calculate the value of the z-statistic or the p-value; (4) decide whether or not to reject H_0; (5) justify your answer and interpret the result based on reality.

For the third problem situation, the aim is for students to recognize that this is a hypothesis test for the difference between two means, for which the following statement is provided:

- *"Two groups of students used different methods to study statistics. Group A used only textbooks; Group B used AI as support (ChatGPT/Copilot). After a test, the results (out of 20 points)."*

The data to consider when solving the problem is as follows:

Group A: 14, 15, 13, 16, 12, 14, 13, 15, 14, 13
Group B: 16, 17, 15, 18, 16, 17, 16, 19, 17, 18

- *Is there a significant difference between the means of the two groups?*

Instructions given by the teacher: (1) formulate the hypotheses H_0 and H_1; (2) consider the significance level $\alpha = 0.05$; (3) ask AI for a t-test for independent samples (Do the t-test for these two samples); (4) Analyze the p-value; (5) reflect on the impact of using AI on the results.

3.3 Pedagogical approach to the first problem situation - students' sleeping hours

In this phase, the teacher contextualizes the first Problem Situation and addresses the students by saying:

- *Let's explore whether the students at our school sleep less than 7 hours a night.*

To do this, they should review the following key concepts: the difference between H_0 and H_1, the meaning of the p-value and the significance level (α), and the test is one-sided ($\mu < 7$). During the resolution, the teacher moves around the classroom and interacts with the students, actively mediating and making some observations:

- *Have you formulated the hypotheses?*
- *Why is this a one-sided test and not a two-sided test?*
- *Does the p-value you considered make sense in light of the sample mean?*

The teacher began the session by encouraging the students to discuss their strategies and solving methods freely, fostering an atmosphere of collaborative sharing. She then sparked a discussion about the traditional resources available to solve statistical problems. Initially shy, the students began to interact with each other until one of the groups volunteered to present its resolution on the board. The teacher took the opportunity to promote debate, questioning the other groups about the results they had obtained and their interpretations. There were disagreements about the value and meaning of the p-value. Faced with this disagreement, the teacher recorded the observation and invited one of the groups with an alternative answer to share its approach. After explaining the different strategies, the students found it easy to identify the most statistically sound answer, reinforcing the importance of critical group analysis. Afterwards, the teacher gave a guided talk on the use of digital resources to support problem-solving, introducing the concept of AI and discussing its applications in the context of research and mathematical problem-solving.

3.3.1 A practical introduction to the use of AI. The teacher presented two AI tools, ChatGPT and Microsoft Copilot, and explained their potential applications in pedagogy. She demonstrated how to formulate transparent and objective questions, how to interpret and evaluate the answers provided by AI, and how they can support mathematical and statistical reasoning. He conducted a practical example that helped students better understand how the technology works, promoting a critical, ethical, and responsible attitude in its use. The importance of using appropriate keywords when formulating AI questions was also stressed. The lecturer pointed out that vague or poorly structured questions can generate inaccurate or decontextualized answers. For example, instead of asking “lower mean” or “statistical test”, it would be more effective and specific to ask:

- *Can you do a t-test to see if the sample mean is less than 7?*

The teacher also warned of the risks of excessively long and confusing questions, which make it difficult for the AI to understand. She used the following inappropriate wording as an example:

- *“We have a set of data and we want to know if the mean can be considered statistically different from the mean of another school because the students sleep little, and we want to know if this is relevant and what to do with the data..."*

This intervention aimed to help students reflect on clarity and precision in mathematical communication, as well as to utilize AI tools as a support for critical thinking, rather than as a substitute for autonomous reasoning. The students are given some guiding questions to think about, about what to ask the AI:

- *Does AI understand everything at once, or should we divide our question into clear and objective parts?*
- *In statistical terms, how can we make our question clearer?*
- *How does AI identify keywords such as “p-value”, “mean”, “significance" or H_0?*

It is explained to the students that the keywords serve as clues for the AI, allowing the tool to select the appropriate statistical method (t-test, z-test, p-value, etc.) and correctly interpret the desired outcome. The lesson continues with some questions for the students, which serve to direct what they want to get from the AI:

- *Do we want to know if there are significant differences?*
- *Is the mean higher or lower than a certain specified value?*
- *Is the proportion different?*

After the theoretical explanation and the initial example, the teacher returned to the first Problem Situation, illustrating it with the statement:

- *I want to know if the students’ mean number of hours of sleep is less than 7.*

In this context, he reminded students of the importance of using relevant keywords, such as “t-test”, ‘mean’, “less than 7”, and “p-value”. She then challenged the students to consider the most effective way to ask AI questions, supporting them in formulating more transparent and more precise questions. Examples of guiding questions included, *“what is the parameter we are testing?”*, *“what is the null hypothesis (H_0) and the alternative hypothesis (H_1)?"*, *“what is the mean value taking into account H_0?”*, *“What is the significance level (α)?”*, and *“what type of test is most appropriate (t, z, one-sided, two-sided)?”*. To support the organization of thought, the teacher explained how to divide the questions progressively and gave examples of how to do this: *“What is the mean of the sample?”*, *“What is the value of the mean that we are going to test?”*, *“What is the sample size?”*, *“What is the alternative hypothesis?”*. Finally, he presented a well-structured instruction:

- *Run a one-sided t-test on this data to see if the mean is less than 7. Consider $\alpha = 0.05$.*

3.3.2 Group Work and Interactions with AI. Each group then chose a problem situation and, based on the examples discussed, formulated the questions to ask the AI to obtain rigorous statistical answers. During this process, the students discussed the results obtained, with the teacher’s support, who used real-time projection to demonstrate how small changes in the wording of the question, such as switching from a one-sided to a two-sided test, can lead to different conclusions. While the students worked on their problem situations, the teacher circulated among the groups, reading out the hypotheses formulated (H_0 and H_1), checking the choice of the appropriate test (t, z, one-sided, two-sided), and helping to reformulate clearer questions for the AI, when necessary. He also confirmed that the p-values were being interpreted correctly.

3.3.3 Mediation of Difficulties and Collective Discussion. One of the groups showed additional difficulties, prompting the teacher to intervene, asking:

- *Is your alternative hypothesis consistent with the problem question?*
- *Does the p-value you obtained indicate evidence against H_0? Why?*

After completing the tasks, the class held a collective discussion based on the projected answers. To encourage reflection, the teacher asked questions such as:

- *What was the result of your sample? Was the mean less than 7?*
- *What was the value of the t-statistic and the p-value?*
- *What decision did you make? Did you reject H_0 or not?*
- *Do you think AI helped to better interpret the problem? Why?*

The teacher then projected an answer generated by the AI, previously selected as clear (or confusing), asking the students to assess its validity. One of the groups compared the AI answer with their own, concluding that they preferred their resolution because it was simpler and they understood the reasoning better. The teacher took the opportunity to emphasize that AI does not replace human statistical reasoning, but only supports it. To deepen the assessment of statistical understanding, the teacher issued a provocative challenge:

- *If the AI told you not to reject H_0 with a p-value of 0.02, what would you say?*

The group answered correctly, "if the p-value is 0.02 and the significance level is 0.05, then as $0.02 < 0.05$, we must reject H_0. There is sufficient evidence against H_0." The teacher continued to stimulate critical thinking with new questions:

- *What if the significance level was 1%?*
- *What if the sample had 50 students?*

The answers given by the group revealed a solid understanding: “If $\alpha = 0.01$, then $0.02 > 0.01$, so we don’t reject H_0”; “with 50 students, the test would be more accurate. With more data, it becomes easier to determine if there is a real difference in sleeping hours.”

3.3.4 Discussion on statistical errors and AI limitations. To assess understanding of type I and II errors, the teacher made the following comment:

- *In the problem situation of hours of sleep, if H_0 is true but the p-value is 0.03 and we reject H_0 with $\alpha = 0.05$, what kind of mistake have we made?*

Responses generated by AI for each group:

- Group I: Type II error, because we rejected H_0 even though it was true (they used ChatGPT).
- Group II: It could be a type I or II error, depending on the interpretation (they used ChatGPT).
- Group III: Type I error, because we rejected H_0 when it is true (they used Microsoft Copilot).

The teacher projected the three answers and asked the groups to evaluate them. Group III acknowledged that only their answer was correct. Group II insisted that theirs also made sense, but eventually recognized that the AI shouldn't give contradictory answers. Group I remained undecided. The teacher took the opportunity to explain that sometimes AI can present "hallucinations" or incorrect answers. She again reminded the students that a Type I error consists of rejecting H_0 when it is true and that a Type II error corresponds to not rejecting H_0 when it is false. Finally, it reinforced the importance of critical thinking, emphasizing that students with essential thinking skills can explain their reasoning, explore multiple resolution strategies, and critique fallacious arguments [16].

3.4 Assessment and Reflection on Learning with AI Support

The aim of learning assessment in this pedagogical practice was not only to verify the acquisition of statistical content, but above all to gauge the development of critical thinking and intellectual autonomy among students when interacting with artificial intelligence tools. To this end, a descriptive evaluation summary was drawn up, focusing on four key dimensions of the work carried out by the students in the context of solving statistical inference problems:

A *Formulating statistical hypotheses - Aims to assess the ability to distinguish and correctly state the null hypothesis (H_0) and alternative hypothesis (H_1), as well as the choice of the appropriate statistical test.*

B *Interpretation of Results - Observes the understanding of the p-value, significance level, and type I and II errors, as well as the coherent justification of the decisions made.*

C *Use of Artificial Intelligence - Analyzes how students used ChatGPT and Copilot as tools for reasoning and supporting statistical thinking, distinguishing between passive and critical use.*

D *Collaboration - Considers students' involvement in group discussions, their ability to explain mathematical strategies, and debate the answers generated using AI.*

This synthesis has been structured into three performance levels: beginner, intermediate, and advanced, allowing for continuous and reflective formative assessment. This approach aims to promote an assessment culture that is more in line with the cognitive demands of the digital age, where in-depth understanding and the ability to question are more valued than simply reproducing procedures. Table 1 shows the criteria defined above.

Table 1: Exploring mathematical strategies for hypothesis testing with AI Criterion Beginner Intermediate Advanced Hypothesis Formulation Struggles to distinguish H_0 and H_1 or formulates incorrect hypotheses. Needs constant support to identify the appropriate type of test. Correctly formulates H_0 and H_1 in simple cases and identifies the test type in most situations, with the occasional backing. Clearly and independently formulates H_0 and H_1, adapting them accurately to the problem context. Selects the appropriate test with confidence. Interpretation of Results Has difficulty interpreting the p-value and significance level. Often makes incorrect decisions. Usually interprets the p-value correctly and makes coherent decisions based on the data, with few errors. Fully understands the meaning of the p-value and type I error. Justifies decisions well and discusses statistical implications clearly. Use of AI Uses AI in a limited way, reproducing responses without understanding. Needs support to interpret outputs. Uses AI functionally, verifying responses and reformulating questions when needed. Learns with occasional support. Uses AI independently, formulates clear statistical questions, critically evaluates results, and integrates AI as a reasoning support tool. Collaboration: Participates minimally in group discussions, with difficulty explaining or justifying the methods used. Contributes to discussions, clearly explaining strategies most of the time, but struggles with more complex justifications. Actively participates in group discussions, clearly explaining and justifying strategies, and encouraging idea exchange with peers.

4 Final Considerations

It was found that the students developed greater autonomy in problem solving, strengthened their mathematical reasoning skills, and deepened their critical awareness of the use of AI. There was an improvement in question formulation skills and in the choice of appropriate strategies, which promoted the choice of more meaningful learning strategies in line with the challenges of using AI in the classroom. These results were in line with research by Zhou et al. [28], in which they examined the influence of generative artificial intelligence on students' critical thinking and problem-solving skills in higher education, finding that the ease of use of AI significantly improves self-regulation, which positively impacts critical thinking and problem-solving skills. They also support the importance, as argued by Trikoili et al. [27], of combining human assessment and AI-based assessment, capitalizing on the strengths of both methods. The pedagogical practice developed highlights the transformative potential of artificial intelligence in promoting critical thinking in math classes. Indeed, as argued in [20], AI has the potential to enhance critical thinking skills in educational environments; however, its implementation must be carefully planned to address both ethical and technical challenges. Ethical and technical difficulties. The integration of AI systems into educational practice can pave the way for more personalized and compelling learning experiences, ultimately benefiting both teachers and students in achieving academic success and boosting intellectual development [2].

By integrating tools such as ChatGPT and Microsoft Copilot into statistical inference tasks, students were challenged to formulate hypotheses rigorously, interpret results based on evidence, and reflect critically on the role of AI in their reasoning. There was an increase in student autonomy, clarity of statistical statements, and the ability to identify errors or ambiguities in answers generated by AI, which are relevant indicators of critical thinking. This experience reinforces the importance of training teachers in the pedagogical use of AI-based tools, not as substitutes for human thought, but as mediators of mathematical dialogue and the development of statistical literacy. Teachers must integrate strategies that promote self-regulation into AI-enhanced learning environments to maximize their impact on student learning (Satone et al., 2025). We recommend continuing exploratory and comparative research into the impact of AI at different levels of education, as well as the creation of curricular references that incorporate critical thinking as a transversal competence in the teaching of mathematics.

Acknowledgements

This work is funded under the Enhancing Inclusive Education in Higher Education Institutions with Artificial Intelligence (AI-Enable) project (2022-1-SI01-KA220-HED000088368), co-funded by the Erasmus+ Programme of the European Union and national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications), and UID/05064/2023 (Valoriza).

References

[1] Salim B Al Maqbali and Nooritawati Md Tahir. 2024. Ai opportunities and risks for students' decision-making skill. In *2024 IEEE 14th International Conference on Control System, Computing and Engineering (ICCSCE)*. IEEE, 321–326.

[2] Ahmad Al Yakin, Ahmed J Obaid, Eka Apriani, Souvic Ganguli, Abdul Latief, et al. 2024. The efficiency of blending ai technology to enhance behavior intention and critical thinking in higher education. In *Embedded Devices and Internet of Things*. CRC Press, 242–266. doi: 10.1201/9781003510420-14.

[3] Robert C. Bogdan and Sari Knopp Biklen. 2003. *Investigação Qualitativa Em Educação: Uma Introdução À Teoria e aos métodos*. Porto.

[4] Borba. 2005. *Humans-with-media and the reorganization of Mathematical Thinking, vol. 39*. Springer US. isbn: 0-387-24264-3.

[5] Lijia Chen, Pingping Chen, and Zhijian Lin. 2020. Artificial intelligence in education: a review. *IEEE Access*, 8, 75264–75278. doi: 10.1109/ACCESS.2020.2988510.

[6] Ann Marcus-Quinn and Tríona Hourigan, editors. 2017. *Transforming mathematics teaching with digital technologies: a community of practice perspective. Handbook on Digital Learning for K-12 Schools*. Springer International Publishing, Cham, 45–57. isbn: 978-3-319-33808-8. doi: 10.1007/978-3-319-33808-8_4.

[7] Sérgio D Correia, Ana Cunha, Maja Pusnik, and Bostjan Sumak. 2024. A micro-learning units package for improving inclusive digital education in hei. In *2024 Digital Inclusion in the Information Society (DIGIN)*. Information Society.

[8] John W. Creswell. 2017. *Qualitative Inquiry and Research Design (international student edition). choosing among five approaches*. SAGE Publications.

[9] Cristina Dias, Carla Santos, Isabel Borges, and Hermelinda Carlos. 2024. A literacia das ferramentas web 2.0 é relevante no contexto educacional? *Latin American Journal of Development*, 6, 1, 360–372. doi: 10.46814/lajdv6n1-025.

[10] Yuanpeng Duan and Chengping Dong. 2023. Risks and countermeasures of artificial intelligence for college students' growth. In *2023 2nd International Conference on Artificial Intelligence and Computer Information Technology (AICIT)*. IEEE, 1–4. doi: 10.1109/AICIT59054.2023.10277729.

[11] PA Facione. 2020. *Critical thinking: What it is and why it counts*. The California Academic Press.

[12] Luciano Floridi. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press, (Aug. 2023). isbn: 978-01988830.98. doi: 10.1093/oso/9780198883098.001.0001.

[13] Sandra Fabijanić Gagro. 2024. Artificial intelligence in education–current challenges. *Annals of the Faculty of Law in Belgrade*, 72, 4, 725–747. doi: 10.51204/Anali_PFBU_24405A.

[14] Joan B. Garfield and Dani Ben-Zvi. 2009. Developing students' statistical reasoning: connecting research and teaching practice. *International Journal of Human–Computer Interaction*, 393–395. doi: 10.1007/s11858-009-0176-6.

[15] Wayne Holmes, Maya Bialik, and Charles Fadel. 2019. *Artificial Intelligence in Education. Promise and Implications for Teaching and Learning*. (Mar. 2019). isbn: 978-1794293700.

[16] Jane Swafford Jeremy Kilpatrick and Bradford Findell. 2001. Adding it up: helping children learn mathematics. *National Academies Press*, (Nov. 2001). doi: 10.17226/9822.

[17] Rose Luckin, Wayne Holmes, Mark Griffiths, and Laurie B. Corcier. 2016. *Intelligence Unleashed: An argument for AI in Education*. Pearson.

[18] Melanie Mitchell. 2019. *Artificial Intelligence: A guide for thinking humans*. Farrar, Straus and Giroux.

[19] Jorge Ortega-Moody, Kouroush Jenab, Saeid Moslehpour, Lizeth Del Carmen Molina Acosta, Edward Jhohan Marin Garcia, and José Neftalí Torres Marin. 2025. Swot analysis of artificial intelligence in education. In *2025 IEEE Engineering Education World Conference (EDUNINE)*. IEEE, 1–6. doi: 10.1109/EDUNINE62377.2025.10981366.

[20] Teresa Chara-De lo Rios, Beymar Solis-Trujillo, Jhon Perez-Ruiz, and María Aquije-Mansilla. 2025. Systematic review of critical thinking using artificial intelligence. *Edelweiss Applied Science and Technology*, 9, 3, (Mar. 2025), 990–1001. doi: 10.55214/25768484.v9i3.5405.

[21] Inayath Ahamed SB et al. 2024. Artificial intelligence effects on student learning outcomes in higher education. In *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*. IEEE, 1–5. doi: 10.1109/ICONSTEM60960.2024.10568868.

[22] Neil Selwyn. 2019. *Should robots replace teachers?: AI and the Future of Education*. English. (1st ed.). Polity Press, United Kingdom, (Nov. 2019). isbn: 978-150952896-7.

[23] Ole Skovsmose. 2011. *An invitation to critical mathematics education*. Ole Skovsmose, editor. SensePublishers. doi: 10.1007/978-94-6091-442-3.

[24] Lubica Stuchlikova and Martin Weis. 2024. From information to insight: reimagining critical thinking pedagogy in the age of artificial intelligence. In *2024 International Conference on Emerging eLearning Technologies and Applications (ICETA)*. IEEE, 591–598.

[25] Boštjan Šumak, Sergio Duarte Correia, Ana Cunha, Tuncer Can, Irfan Simsek, Katja Kous, and Maja Pušnik. 2024. Identification of factors that impact e-inclusion in hei. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*. IEEE, 478–483. doi: 10.1109/MIPRO60963.2024.10569746.

[26] Boštjan Šumak et al. 2024. Ai-based education tools for enabling inclusive education: challenges and benefits. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*. IEEE, 472–477. doi: 10.1109/MIPRO60963.2024.10569714.

[27] Anna Trikoili, Despoina Georgiou, Christina Ioanna Pappa, and Daniel Pittich. 2025. Critical thinking assessment in higher education: a mixed-methods comparative analysis of ai and human evaluator. *International Journal of Human–Computer Interaction*, 1–14. doi: 10.1080/10447318.2025.2499164.

[28] Xue Zhou, Da Teng, and Hosam Al-Samarraie. 2024. The mediating role of generative ai self-regulation on students' critical thinking and problem-solving. *Education Sciences*. doi: 10.3390/educsci14121302.

[29] Xiaozhu Zou, Ping Su, Lexing Li, and Ping Fu. 2024. Ai-generated content tools and students' critical thinking: insights from a chinese university. *IFLA journal*, 50, 2, 228–241. doi: 10.1177/03400352231214963.