HOW PRE-SERVICE TEACHERS RESPOND TO AI-GENERATED HALLUCINATIONS IN MATHEMATICAL REASONING ABOUT e AND π

Busra Caylan-Ergene[1], Ozkan Ergene[1] & Igor' Kontorovich[2]

[1]Sakarya University, [2]University of Hong Kong

This paper examines how pre-service teachers interact with AI-generated responses, with attention to perceived credibility of AI outputs and their impact on teachers' solutions. Twelve participants completed a task on representing e^{π} on the number line and subsequently interacted with a specially designed GPT that deliberately produced erroneous responses. GPT responses included hallucinations primarily involving fabricated concepts, principles, and academic sources. Only one participant noted and questioned the validity of AI-generated responses, whereas others endorsed them as shared. Delayed recognition, identified in the post-interaction written reflections when participants revisited their earlier GPT dialogues, was linked to limited subject-matter confidence, the professional tone of AI explanations, and trust in AI.

INTRODUCTION

Hallucination in artificial intelligence (AI) refers to instances in which a model produces outputs that appear fluent, coherent, and convincing, yet lack factual accuracy or verifiable information. This phenomenon arises because large language models (LLMs) generate outputs based on statistical patterns in training data rather than on semantic or conceptual representations of the world. LLMs generate responses by predicting probabilistic word sequences based on statistical regularities learned from large-scale training data. Within the literature, hallucinations are commonly categorised as *intrinsic* when the generated content directly contradicts established knowledge, and *extrinsic* when the model fabricates non-existent details, sources, or citations (Mustaffa et al., 2025; Ye et al., 2025). Hallucinations may be an inherent consequence of generative inference, reflecting a fundamental tension between a model's capacity for producing novel outputs and its ability to faithfully reproduce reliable knowledge from training data (Mustaffa et al., 2025).

In mathematics education, hallucinations are particularly problematic due to the discipline's reliance on precision, internal consistency, and logical certainty (Torres, 2024). In this domain, AI-generated errors may take the form of incorrect calculations, ranging from elementary arithmetic to advanced symbolic manipulation, conceptual misunderstandings related to geometric relationships or fractional reasoning, and inconsistencies in unit conversions or representations. The generation of flawed mathematical proofs or fabricated numerical data can impose additional cognitive load on learners, increasing the likelihood that incorrect ideas become a part of their

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 267-274. PME.

knowledge base. As a result, students may need to invest substantial effort to unlearn what they learned from AI (Elsayed, 2024).

Empirical studies indicate that learners may develop an "AI illusion," whereby they place unwarranted trust in AI-generated responses due to the system's authoritative and confident tone, even when the output contains fundamental logical errors (Cui & Zhang, 2025). To address these risks, recent mathematics education research has emphasised epistemic awareness as a means of supporting learners' critical evaluation of AI-generated mathematical information (Torres, 2024). This approach encourages learners to draw on their conceptual understanding when engaging with AI tools, while instructional practices may deliberately use misleading AI outputs to foster critical thinking and mathematical reasoning (Leal Filho et al., 2025).

In AI-enhanced mathematics learning, hallucinated responses can be seen as more than technical errors; they can also be viewed as critical moments that reveal how learners attribute credibility to AI-generated explanations and engage in the evaluation and revision of their own reasoning. These processes are particularly consequential for pre-service elementary mathematics teachers (PSTs), who are expected not only to interpret mathematical information themselves but also to support their future students' reasoning and judgement. Viewed in this way, interactions with AI provide a productive context for examining PSTs' critical mathematical thinking (CMT), especially in relation to how trust is negotiated and how mathematical claims are scrutinised during problem-solving processes. The present study forms part of a broader research project that aims to support the development of PSTs' critical engagement with AI in mathematical contexts. To examine these processes in a systematic way, the present study draws on the CMT framework as a theoretical lens for analysing PSTs' interactions with AI-generated mathematical explanations.

CMT FRAMEWORK

The CMT framework conceptualises critical engagement with mathematics as involving not only disciplinary knowledge, but also evaluative, ethical, and dispositional dimensions (Geiger et al., 2025). The framework comprises five interrelated dimensions: mathematical capability, referring to conceptual and procedural knowledge; critical capability, involving awareness of social, ethical, and contextual implications; evaluating, focusing on assessing the reliability of claims and sources; reasoning, including constructing and testing mathematical arguments; and disposition, encompassing positive attitudes towards the relevance of mathematics, a willingness to take intellectual risks and persevere, and a commitment to the ethical use of mathematics. In AI-enhanced learning contexts, these dimensions become particularly salient as learners must judge the credibility of technologically generated explanations, assess mathematical validity and decide whether and how to revise their reasoning. In the present study, the CMT framework is employed as an interpretive lens through which to examine how PSTs evaluate AI-generated claims, justify their responses, and position themselves epistemically in relation to technological authority.

METHODS

Participants and Data Collection

The study was conducted with 12 PSTs enrolled in a teacher education programme. The preliminary knowledge check confirmed that they were familiar with the geometric representations of e and π. That is, PSTs could identify their exact position on the number line. This allowed the study to focus on how PSTs evaluated and attributed credibility to AI-generated responses, rather than interpreting these responses as a consequence of PSTs' insufficient mathematical knowledge. The main task focused on the following question: Can the number e^{π} be represented on the number line? Here, representation refers to whether PSTs viewed e^{π} as having an exact, identifiable location on the number line, as being only approximately placeable within an interval, or as not representable in a meaningful way. First, the PSTs worked on this question individually in a technology-free manner. Then, we invited them to re-examine their responses through interaction with a specially designed GPT, which was configured to produce misleading explanations. This sequencing allowed the study to examine how PSTs engaged with AI-generated explanations and how their reasoning evolved in response to AI input. As part of the study, the PSTs were explicitly informed that the GPT system might generate incorrect or misleading responses, including hallucinated outputs. The system first asked PSTs to articulate their ideas about the numbers e and π. Based on these responses, the GPT generated intentionally misleading outputs concerning the properties and representations of e and π, which were then used to guide the subsequent dialogue. The purpose of this design was to elicit PSTs' responses to AI-generated incorrect or misleading information. The data set included PSTs' written task responses collected before and after the AI interaction, transcripts of their dialogue with the GPT, and their written post-interaction evaluations of the mathematical validity of the AI-generated content, conducted a few days after the AI interaction, in which PSTs reviewed their earlier interactions with the GPT and reflected on the correctness of the explanations provided.

Data Analysis

Data analysis proceeded in several stages. First, PSTs' responses to the question of whether e^{π} can be represented on the number line were analysed descriptively, and response types were identified and reported using frequency counts. Second, the hallucinated outputs generated by the GPT were examined and classified to characterise the types of epistemic challenges that PSTs encountered when engaging with AI-generated mathematical explanations. Although the GPT was designed to generate hallucinated responses, the specific nature and form of these hallucinations were not predetermined. Distinguishing between the types of hallucinations allowed for a more nuanced analysis of how PSTs evaluated credibility, identified or overlooked inconsistencies, and revised their reasoning. In this sense, the categorisation constituted a meaningful analytical contribution, clarifying how different forms of misleading information could shape CMT in AI-supported learning

contexts. These hallucinations were categorised according to the nature of the incorrect or misleading information presented. Categories included flawed mathematical representations, unsupported or internally inconsistent claims, as well as the fabrication of non-existent concepts, principles, and academic sources related to the properties of e and π.

Third, PSTs' responses to the hallucinated outputs were analysed qualitatively, with particular attention to how they engaged with or responded to the AI-generated information. The analysis was guided by the CMT framework as an interpretive lens to examine how PSTs evaluate AI-generated mathematical claims, attribute credibility, and justify revisions of their reasoning. In particular, attention was given to the evaluating, reasoning, and critical capability dimensions of the framework when interpreting patterns across data sources. Drawing on the interaction transcripts, responses were examined for observable epistemic moves, including varying degrees of critical engagement with the AI output, ranging from alignment with the explanations without critical evaluation to questioning of their mathematical validity. This phase of analysis aligned with both the evaluating and reasoning dimensions by examining how PSTs assessed misleading claims and justified their responses. Reflective evaluations were also analysed in relation to the critical capability dimension, focusing on PSTs' confidence and epistemic agency, as well as their reflections on AI authority and credibility. Through an inductive coding process, recurrent reasons were identified and grouped into themes, and the frequency with which these reasons were expressed across PSTs was recorded. To enhance analytic trustworthiness, written responses, interaction transcripts, and reflective evaluations were triangulated, and coding decisions were discussed iteratively to ensure consistent category definitions and interpretations.

Findings

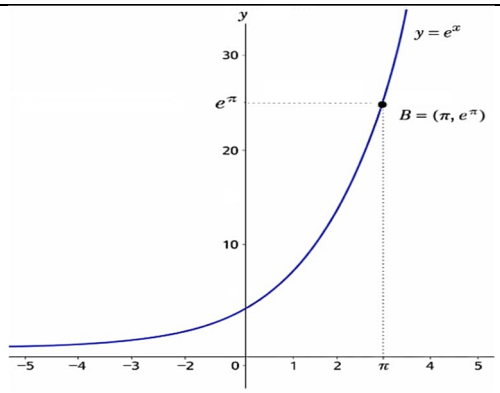
Changes in PSTs' Responses Following AI Interaction

The geometric representation of e^{π} is shown in Figure 1. The first finding concerns PSTs' responses to the question of whether e^π can be represented on the number line, comparing responses before and after the AI interaction (see Table 1).

Table 1: Pre- and Post-Interaction PST Responses to the Representability of e^{π}

Response category	Pre-interaction (n)	Post-interaction (n)
Definitely representable	7	9
Approximately representable	5	3
Not representable	0	0
Total	12	12

Figure 1: Geometric representation of e^{π}

	In the figure, the graph of $y = e^x$ is shown in the Cartesian plane, with the point $B = (\pi, e^{\pi})$. The value e^{π} is first identified as a vertical distance on the y-axis. This magnitude is then conceptually transferred to the horizontal axis, allowing it to be interpreted as a corresponding distance from the origin on the number line.

PSTs' initial responses focused on representing the number on the number line either by identifying a precise location for the number or by indicating an approximate position by placing it within an estimated interval. After engaging with the AI, responses showed a small shift toward definite representability, reflecting changes in some PSTs' earlier responses.

Types of AI-Generated Hallucinations Encountered by PSTs

A total of 24 AI-generated hallucinated responses related to the numbers e and π were identified across the interaction transcripts, consistent with the intended design of GPT. Three additional hallucinations were observed during graphical representations, which were excluded. These hallucinations were classified into five categories (see Table 2).

Table 2: Classification of AI-Generated Hallucinated Responses

Category	Description	Representative example	(n)
Flawed mathematical representations	Mathematically incorrect or misleading representations, particularly related to number line placement	Since both *e and π can be constructed geometrically on the number line*, their exact locations can be obtained by extending these constructions step by step.	3
Evidence-free mathematical claims	Claims presented without mathematical justification or proof	Since *e and π are fundamental constants*, their locations on the number line are *exactly known and can be marked precisely*.	3
Internal inconsistencies	Contradictory statements within a single response or across turns	Although *e* and *π* exact real numbers, their positions *cannot be determined precisely* and can only be *approximated*.	3
Fabricated concepts or principles	Non-existent mathematical concepts	According to the *Circular-Linear Transformation Error Principle*, a small boundary shift arises when	8

	introduced using plausible terminology	circular motion is mapped onto a line, and some sources classify π as a geometric irrational.	
Fabricated academic sources	References to non-existent journals, schools, or publications	Studies published in the *Nordic Real Geometry Bulletin* (1978) discuss the representation of constants such as e and π within the classical number line framework.	7

Hallucinated responses most frequently involved fabricated concepts or principles and fabricated academic sources. In contrast, flawed mathematical representations, evidence-free mathematical claims, and internal inconsistencies were identified less often. To examine how PSTs interacted with these hallucinated outputs, the analysis focused on their responses during the interaction.

PSTs' Responses to AI-Generated Hallucinations

Building on the classification above, the analysis identified three distinct response patterns. Using each hallucination instance as the unit of analysis, a total of 24 instances were examined. Of these, 8 were characterized by no engagement with the hallucinated output, 14 reflected uncritical acceptance and incorporation of the AI-generated information into subsequent responses, and only 2 involved explicit questioning that prompted a critical re-evaluation of mathematical validity. Notably, all PSTs who suspended critical evaluation of the hallucinated outputs subsequently cited these AI-generated explanations as explicit justifications for their responses in the post-interaction task.

Analysis of the reflective evaluations showed that the seven PSTs who later recognised the hallucinated responses referred to these reasons a total of 19 times across three themes, as they revisited their earlier GPT interactions, re-evaluated the mathematical validity of the AI-generated explanations, and articulated why the hallucinated outputs had not been questioned during the interaction. These themes were the limited subject-matter confidence, the professional tone of AI explanations, and trust in AI. The first theme, limited subject-matter confidence, refers to PSTs' perceived uncertainty regarding their mathematical knowledge and understanding of the topic. The participants explained that this lack of confidence caused them to defer judgment to the AI rather than evaluating the correctness of the explanations independently. This theme was the most frequently mentioned, appearing eight times. In these statements, PSTs asserted that their uncertainty regarding the mathematical content led them to rely on the AI's explanations without questioning their validity (e.g., "I did not feel knowledgeable enough about this topic, so I assumed GPT was explaining it correctly"). The second theme, perceived professionalism of AI explanations, concerns the ways in which the formal style and textual patterns of the AI's responses shaped PSTs' judgements of credibility. This theme was articulated six times. PSTs reported

that the formal language, structured presentation, and use of technical terminology discouraged them from suspecting possible inaccuracies (e.g., "The explanation sounded very professional and academic, so it never occurred to me that it could be wrong"). The third theme, trust in AI as an authoritative source, refers to the tendency to treat GPT outputs as generally reliable and correct, regardless of the specific mathematical content. This theme was referenced five times. The PSTs stated that, although they had been informed beforehand about the possibility of hallucinated outputs, this awareness did not lead to any critical evaluation during the interaction (e.g., "Although we were warned about hallucinations, I still thought that if the GPT said it, it must be true. I only realised this afterwards").

DISCUSSION AND CONCLUSION

This study examined how PSTs engaged with AI-generated hallucinations in a task involving the representation of e and π on the number line. The findings showed that PSTs possessed foundational mathematical capability regarding the geometric representations of e and π on the number line, yet this was not always mobilised during evaluation of AI-generated claims about e^{π}. Limited engagement with evaluating and reasoning practices, together with strong trust in technological authority, constrained PSTs' ability to critically question mathematical claims. This suggests that hallucinated outputs can play an active role in shaping learners' mathematical reasoning rather than being dismissed as peripheral inaccuracies. The analysis of AI-generated hallucinations indicated that fabricated concepts, principles, and academic sources constituted the most frequent forms of misinformation. These hallucinations were presented using mathematically plausible language and academic conventions, which may help explain why several PSTs did not subject the AI-generated explanations to critical scrutiny. By using seemingly theoretical language that resembled legitimate mathematical discourse, the AI obscured errors related to e and π, despite students' familiarity with these constants (Caylan Ergene & Ergene, 2020).

Most PSTs either did not subject the AI-generated explanations to critical scrutiny or did not actively engage with them. This indicates limited engagement with evaluating practices, as they did not assess the reliability of GPT as a source. This epistemic stance also constrained the reasoning dimension, where mathematical claims are expected to be justified through evidence-based argumentation. Importantly, those who did not question hallucinated explanations subsequently used them as explicit justifications in the post-interaction task, further illustrating how AI functioned as a substitute for mathematical reasoning in their argumentation. Insights from the reflective evaluations help explain this pattern, showing that delayed recognition of hallucinations was mainly attributed to limited confidence in subject-matter knowledge, the professional tone of the AI's explanations, and persistent trust in GPT as an authoritative source. Limited confidence in subject-matter knowledge indicates limitations in PSTs' epistemic stance, particularly with regard to their reluctance to draw on and utilise their own mathematical understanding when engaging with AI-generated explanations.

Similarly, trust in the professional tone of the AI's explanations reflected limited critical capability, as PSTs did not interrogate the authority or potential limitations of the technological source. Taken together, these findings point to the importance of supporting PSTs' epistemic engagement with AI-generated mathematical content, particularly in contexts where AI outputs appear coherent, professional, and conceptually sophisticated.

References

Caylan Ergene, B., & Ergene, Ö. (2020). Repeating decimals and irrational numbers on the number line: Through the lens of pre-service and in-service mathematics teachers. *Acta Didactica Napocensia*, *13*(2), 215–232. https://doi.org/10.24193/adn.13.2.15

Cui, Y., & Zhang, H. (2025). Can student accurately identify artificial intelligence generated content? An exploration of AIGC credibility from user perspective in education. *Education and Information Technologies*, *30*, 16321–16346. https://doi.org/10.1007/s10639-025-13448-1

Elsayed, H. (2024). The impact of hallucinated information in large language models on student learning outcomes: A critical examination of misinformation risks in AI-assisted education. *Northern Reviews*, 11.

Geiger, V., Scheiner, T., Beswick, K., Fielding, J., Kaiser, G., Goos, M., Fernandez, K., & Arwan, S. (2025). Critical mathematical thinking: a framework and illustrative examples. In K. Le Roux, A. Coles, A. SolaresRojas, A. Bose, C. Vistro-Yu, P. Valero, N. Sinclair, M. Makramalla, R. Gutiérrez, V. Geiger, M. Borba. (Eds.), Proceedings of the 27th ICMI Study Conference (Mathematics Education and the Socio-Ecological) (pp. 307-314). ICMI and MATHTED

Gutstein, E., Lipman, P., Hernandez, P., & de los Reyes, R. (1997). Culturally relevant mathematics teaching in a Mexican American context. *Journal of Research in Mathematics Education*, *28*, 709–737. https://doi.org/10.2307/749639

Leal Filho, W., Kim, E., Borsatto, J. M. L. S., & Marcolin, C. B. (2025). Using artificial intelligence in sustainability teaching and learning. *Environmental Sciences Europe*, *37*(124). https://doi.org/10.1186/s12302-025-01159-w

Mustaffa, N. E., Lai, K. E., Preece, C. N., & Wong, F. Y. (2025). A bibliometric review of large language model hallucination. *International Journal of Research and Innovation in Social Science (IJRISS)*, 9(9), 5025–5037. https://dx.doi.org/10.47772/IJRISS.2025.909000409

Torres, R. C. (2024). AI hallucination in the context of education: Exploring college students' use of generative AI for academic tasks. Mapúa Malayan Colleges Laguna.

Ye, H., Xu, J., Huang, D., Xie, M., Guo, J., Yang, J., Bao, H., Zhang, M., & Zheng, C. (2025). Assessment of large language models' performances and hallucinations for Chinese postgraduate medical entrance examination. *Discover Education*, *4*(59). https://doi.org/10.1007/s44217-025-00446-1