

The Prisoner's Dilemma of AI: Cognitive Offloading Among Future Mathematics Teachers

Otto Suchanek, Tomas Javorcik, Tatiana Havlaskova and Magdalena Zavodna

Faculty of Education, University of Ostrava, Czech Republic

otto.suchanek@osu.cz

tomas.javorcik@osu.cz

tatiana.havlaskova@osu.cz

magdalena.zavodna@osu.cz

Abstract: The integration of Large Language Models (LLMs) into education has introduced substantial opportunities as well as challenges related to students' cognitive processes and professional development. For pre-service teachers, generative Artificial Intelligence (AI) creates a professional paradox. While AI systems can increase efficiency, support learning processes, and provide rapid access to information, excessive reliance on such systems may lead to cognitive offloading, automation bias, and gradual weakening of critical thinking and professional autonomy. Within educational contexts, this tension may be interpreted as a conflict between short-term individual gains and long-term consequences for human expertise and independent reasoning. This study compares pretest and posttest perceptions of generative AI among first-year pre-service mathematics teachers following a targeted educational intervention focused on principles of AI functioning and related cognitive risks. The Prisoner's Dilemma is used as a conceptual and pedagogical lens for framing the tension between the immediate benefits of AI use and its possible long-term educational consequences; it is not empirically operationalised or tested as a game-theoretic model. Within this framework, delegating cognitive activities to AI may represent a rational strategy at the individual level while simultaneously producing undesirable outcomes for broader educational development. A quasi-experimental pretest–posttest design without a control group was employed. Participants consisted of undergraduate students enrolled in a Mathematics Teacher Education programme at the University of Ostrava (pretest $n = 18$; posttest $n = 14$). Data collection was conducted using a semantic differential instrument consisting of sixteen bipolar dimensions measuring students' perceptions of AI characteristics. The intervention included instructional activities explaining principles of large language models, cognitive offloading, automation bias, and ethical implications of excessive dependence on AI systems. Data were analysed using descriptive statistics and non-parametric methods, specifically the Mann–Whitney U test. The results did not reveal statistically significant differences between pretest and posttest measurements across any of the analysed dimensions. Several mean scores differed descriptively between the measurements, particularly for transparency, predictability, and the role of AI in deeper learning, but these differences were not statistically significant. Given the small, unmatched samples and the absence of a control group, they cannot be interpreted as evidence of an intervention effect. These results highlight the importance of integrating critical AI literacy into teacher education programmes.

Keywords: Artificial Intelligence, Cognitive Offloading, Game Theory, Mathematics Education, Pre-service Teachers

1. Introduction

The integration of large language models (LLMs) into education represents one of the most significant transformations in the contemporary educational environment. Unlike previous technological innovations, tools based on generative artificial intelligence, such as ChatGPT, Gemini, and Claude, have spread among students and teachers more rapidly than stable institutional, methodological, and ethical frameworks for their use have been established. As a result, students and educators now commonly work with systems capable of generating texts, solving tasks, and creating educational materials without necessarily possessing sufficient understanding of the underlying principles and limitations of these models (Kasneci et al, 2023).

From a pedagogical perspective, LLMs do not represent merely another digital tool but rather a substantial transformation of the structure of the learning process itself. Generative models alter the ways in which information is searched, knowledge is organised, and problems are solved. At the same time, part of the cognitive work is transferred from humans to algorithms, raising the issue of *cognitive offloading*, defined as the delegation of mental operations to external systems. Current research suggests that this process may produce a dual effect. On one hand, AI may reduce cognitive load and allow students to focus on higher-order thinking processes; on the other hand, it may weaken deeper understanding, metacognitive control, and long-term intellectual autonomy (Grinschgl et al, 2025; Zhai, Wibowo and Li, 2024).

The issue of metacognitive calibration represents a particularly important challenge in the context of generative AI. Metacognitive calibration refers to an individual's ability to realistically evaluate their own knowledge and competencies. Studies indicate that AI-supported environments may create an illusion of understanding, in which students overestimate their actual abilities because they can rapidly generate plausible-looking answers

without deeper comprehension of the problem. If students repeatedly accept AI outputs without critical verification, their ability to independently monitor their own thinking may gradually decline, increasing reliance on algorithmic support systems (Lee et al, 2025; Lin and Sohail, 2026).

This issue is particularly significant in mathematics education and other exact disciplines. Mathematical reasoning relies on logical precision, structured argumentation, and the ability to identify even minor errors in reasoning processes. Although LLMs can effectively explain procedures or generate examples, and some studies suggest a positive influence on students' conceptual understanding and mathematical confidence (Canonigo, 2024), their operation is fundamentally based on probabilistic prediction of subsequent tokens rather than genuine logical reasoning (Soygazi and Oğuz, 2023). Generative models therefore do not “think” in the human sense but instead produce outputs based on statistical patterns in data. The authoritative nature of such responses may consequently lead students toward uncritical acceptance of incorrect mathematical procedures or arguments.

Closely related to this phenomenon are the concepts of *automation bias* and *automation complacency*. Automation bias refers to the tendency to trust automated systems even in situations where their outputs contradict human judgement or contain errors. Repeated experiences with apparently successful AI systems may reduce the level of critical monitoring and support the development of over-reliance on algorithmic systems. Research also suggests that frequent delegation of analytical tasks to AI may be associated with weakening of certain cognitive processes, particularly in areas requiring deeper logical analysis and independent argumentation (Parasuraman and Manzey, 2010; Nguyen, 2024; Zhai, Wibowo and Li, 2024).

Within the context of preparing future mathematics teachers, an important professional paradox emerges. Generative AI offers substantial improvements in efficiency, speed, and accessibility of supportive materials; however, it may simultaneously weaken the development of competencies that constitute the foundation of teaching professionalism. Current AI literacy frameworks therefore emphasise that teacher education should focus not only on technical mastery of tools but also on the development of critical thinking, epistemological vigilance, and the ability to reflect on the limitations of algorithmic systems (Bilbao Eraña and Arroyo-Sagasta, 2025; Hsu, 2026).

One possible conceptual lens for interpreting this paradox is the Prisoner's Dilemma, used here as a pedagogical analogy rather than as an empirically tested game-theoretic model. From a short-term perspective, delegating part of cognitive work to AI may appear rational because it provides immediate advantages in efficiency and performance. However, if such behaviour becomes the dominant strategy within a group, it may result in a collective loss of deeper understanding, professional autonomy, and critical thinking abilities. Evolutionary models based on the Prisoner's Dilemma further suggest that changes in the “payoff matrix” may substantially influence strategic behaviour and support the emergence of cooperative actions (Perc et al, 2024).

Although current literature increasingly discusses AI literacy among future teachers (Bilbao Eraña and Arroyo-Sagasta, 2025; Hsu, 2026), automation bias and over-reliance on AI systems (Parasuraman and Manzey, 2010; Nguyen, 2024), as well as cognitive offloading in generative AI environments (Grinschgl et al, 2025; Zhai, Wibowo and Li, 2024), relatively limited attention has been devoted to a game-theoretical interpretation of these processes in the context of mathematics teacher education. Existing studies primarily focus on technical AI integration, pedagogical effectiveness, or AI competency development, whereas the strategic conflict between short-term efficiency and long-term professional autonomy remains insufficiently explored (Perc et al, 2024).

Therefore, this study compares pretest and posttest perceptions of generative AI among pre-service mathematics teachers in the context of an educational intervention focused on the principles of LLM functioning, automation bias, cognitive offloading, and game theory. The research employs a pretest–posttest design based on semantic differential scales to analyse differences between the two measurements.

2. Methodology

2.1 Participants and Research Design

The study employed a quasi-experimental pretest–posttest design without a control group. The aim of the research was to compare pretest and posttest perceptions of generative artificial intelligence in the context of a targeted educational intervention focused on the principles of AI functioning, cognitive offloading, automation bias, and the application of game theory within an educational context.

The study used a convenience sample of first-year undergraduate students enrolled in a Mathematics Teacher Education programme at the Faculty of Education, University of Ostrava. A total of 18 students participated in the pretest, while 14 students completed the posttest after the intervention. The difference in the number of respondents was caused by regular absenteeism during face-to-face classes. Participation in the study was voluntary, and all participants were informed about the purpose of the research.

The study was conducted in three consecutive phases:

1. Pretest
2. Educational intervention
3. Posttest

The following research questions were formulated:

RQ1: What differences in perceptions of generative AI are observed between the pretest and posttest measurements among pre-service mathematics teachers?

RQ2: Do the pretest and posttest measurements differ in dimensions related to a critical and realistic perception of generative AI?

The study followed institutional ethical guidelines, and all participants provided informed consent before participation.

2.2 Research Procedure

The educational intervention was implemented between the pretest and posttest and consisted of two interconnected instructional blocks focusing on the principles of generative AI, cognitive offloading, automation bias, and the ethical implications of over-reliance on AI systems.

2.2.1 Block 1: How Large Language Models Work

The first part of the intervention focused on explaining the principles underlying neural networks and large language models. Students were introduced to the idea that neural networks do not represent a “digital mind” but rather mathematical models based on recognising statistical patterns in data.

The instructional content included explanations of:

- next-token prediction,
- attention mechanisms,
- embeddings,
- vector representations of words,
- high-dimensional vector spaces.

To explain vectorisation principles, a food-taste analogy was used. Students were shown that each food item can be described using several characteristics (e.g., sweetness, saltiness, sourness, and bitterness), creating a vector representation.

The analogy was subsequently transferred to language models, where words are represented using hundreds or thousands of dimensions describing semantic relationships between concepts. Students were also introduced to a simplified example:

$$\text{King} - \text{Man} + \text{Woman} = \text{Queen}$$

as an illustration of mathematical operations performed on vector representations of words.

The purpose of this block was to reduce anthropomorphic perceptions of AI and strengthen understanding of the statistical nature of language models.

2.2.2 Block 2: AI as a Tool and a Trap

The second part focused on ethical and pedagogical implications of AI use. Students initially completed a SWOT analysis focused on the use of AI in educational contexts in order to identify potential benefits and risks associated with AI systems.

Algorithmic atrophy was subsequently discussed through the following question:

"Would you be able to identify an error in an AI-generated mathematical proof if you could not construct the proof yourself?"

A key part of the intervention involved applying the Prisoner's Dilemma framework to AI use in educational settings. The Prisoner's Dilemma describes situations in which individually rational decisions may lead to collectively suboptimal outcomes. In this context, excessive reliance on AI was presented as a potentially beneficial short-term strategy that may produce undesirable long-term consequences for learning and professional development.

Students discussed a simulated scenario involving an algebra test with two possible strategies:

- Option A – study independently
- Option B – use hidden AI support

Using concepts of Nash equilibrium and Pareto optimality, students discussed the conflict between short-term individual benefits and long-term consequences for professional development. The payoff matrix represented possible outcomes associated with different strategies, such as independent learning versus excessive use of AI. Students discussed how changes in educational conditions may alter incentives and shift behaviour towards more desirable learning outcomes. In this study, however, the Prisoner's Dilemma served solely as a conceptual and pedagogical framework within the intervention; no payoff parameters, strategy choices, equilibria, or other game-theoretic variables were operationalised or tested in the empirical analysis.

The intervention also included activities designed to modify these incentives. For example:

"AI generated the following mathematical procedure. Identify three logical errors and explain why they occurred."

2.3 Research Instrument and Statistical Analysis

Data collection was conducted using a questionnaire based on a semantic differential approach. The instrument consisted of sixteen bipolar scales measured using a seven-point semantic differential. Higher values represented more positive evaluations of the given characteristic.

Students evaluated AI using dimensions such as:

- Non-transparent – Transparent
- Biased – Impartial
- Unpredictable – Predictable
- Competitive – Cooperative
- Manipulative – Honest
- Reducing understanding – Deepening understanding
- Leading to surface learning – Leading to deep learning
- Passive – Activating
- Dangerous – Safe
- Unethical – Moral

among others.

Due to the ordinal nature of data obtained through semantic differential scales, non-parametric statistical procedures were employed. Data analysis included descriptive statistics (mean, median, and standard deviation) followed by comparisons between pretest and posttest results.

Because pretest and posttest data were not paired at the individual respondent level, paired statistical procedures could not be applied. Therefore, pretest ($n = 18$) and posttest ($n = 14$) responses were analysed as two independent samples using the Mann–Whitney U test.

Arithmetic means were used only for descriptive and graphical purposes. Effect sizes were additionally examined to support interpretation of practical significance. Given the small group sizes, statistical power was limited, particularly for detecting small effects. Accordingly, non-significant results were not treated as evidence of equivalence or of the absence of any difference. All statistical analyses were performed using NCSS software.

3. Results

3.1 Descriptive Statistics

Descriptive statistics including mean (M), standard deviation (SD), median (Md), and interquartile range (IQR) were calculated for both pretest and posttest measurements. Given the ordinal nature of the seven-point scale, medians and interquartile ranges were additionally considered because they provide a more robust description of central tendency and variability.

In the pretest measurement, the highest mean scores were observed for the dimensions *Confrontational–Supportive* (M = 5.17), *Self-oriented–Prosocial* (M = 5.11), and *Reducing Understanding–Deepening Understanding* (M = 4.83), suggesting a relatively positive perception of generative AI prior to the intervention. The lowest scores were found for *Passive–Activating* (M = 3.61), *Unpredictable–Predictable* (M = 3.72), and *Leading to Surface Learning–Leading to Deep Learning* (M = 3.72). Following the intervention, *Confrontational–Supportive* remained the highest-rated dimension (M = 5.43), followed by *Biased–Impartial*, *Competitive–Cooperative*, and *Self-oriented–Prosocial* (all M = 4.71). In contrast, *Leading to Surface Learning–Leading to Deep Learning* (M = 3.21) and *Passive–Activating* (M = 3.79) remained among the lowest-rated dimensions.

Overall, the descriptive results suggest relatively stable perceptions of AI across most dimensions, while several dimensions associated with deeper learning and cognitive engagement showed less favourable evaluations following the educational intervention.

Table 1: Descriptive Statistics for Pretest and Posttest Measurements

Item	Pretest M	Pretest SD	Pretest Md	Pretest IQR	Posttest M	Posttest SD	Posttest Md	Posttest IQR
Non-transparent - Transparent	3.83	1.27	4	1.25	4.21	1.14	4	2
Biased - Impartial	4.61	1.26	4.5	2.25	4.71	1.35	5	2
Unpredictable - Predictable	3.72	1.14	4	1.25	4.21	1.4	4	2.25
Difficult - Easy to Use	4.78	1.02	4.5	2	4.57	1.48	5	3
Competitive - Cooperative	4.44	1.53	4	3	4.71	1.47	5	3
Self-oriented - Prosocial	5.11	1.34	5	2.25	4.71	1.16	5	1.25
Manipulative - Honest	4.5	1.27	5	1.25	4.43	1.31	5	2.25
Confrontational - Supportive	5.17	1.31	5	2	5.43	0.87	5.5	1
Reduces Understanding - Deepens Understanding	4.83	1	5	0.5	4.64	1.42	5	2.25
Surface Learning - Deep Learning	3.72	1.14	3	1.25	3.21	1.28	3	1.25
Discouraging Thinking - Supporting Thinking	3.83	1.45	4	2.25	4.07	1.52	4	2.25
Passive - Activating	3.61	1.26	3	1.25	3.79	1.34	4	2
Unregulated - Regulated	4.11	1.39	4	2	4	1.39	4	2.25
Dangerous - Safe	4.17	1.42	4	2	4.5	0.85	4	2
Harmful - Helping Humans	5.11	1.08	5	2	4.86	1.06	5	2

Item	Pretest M	Pretest SD	Pretest Md	Pretest IQR	Posttest M	Posttest SD	Posttest Md	Posttest IQR
Unethical-Moral	4.56	1.15	5	1	5	1.12	5	2

Note: M = Mean; SD = Standard Deviation; Md = Median; IQR = Interquartile Range.

3.2 Semantic Differential Profile

To visualise changes in the perception of generative artificial intelligence between pretest and posttest measurements, a semantic differential profile was created. The figure presents mean scores for individual dimensions measured on a seven-point scale, where higher values indicate more positive evaluations of a given characteristic.

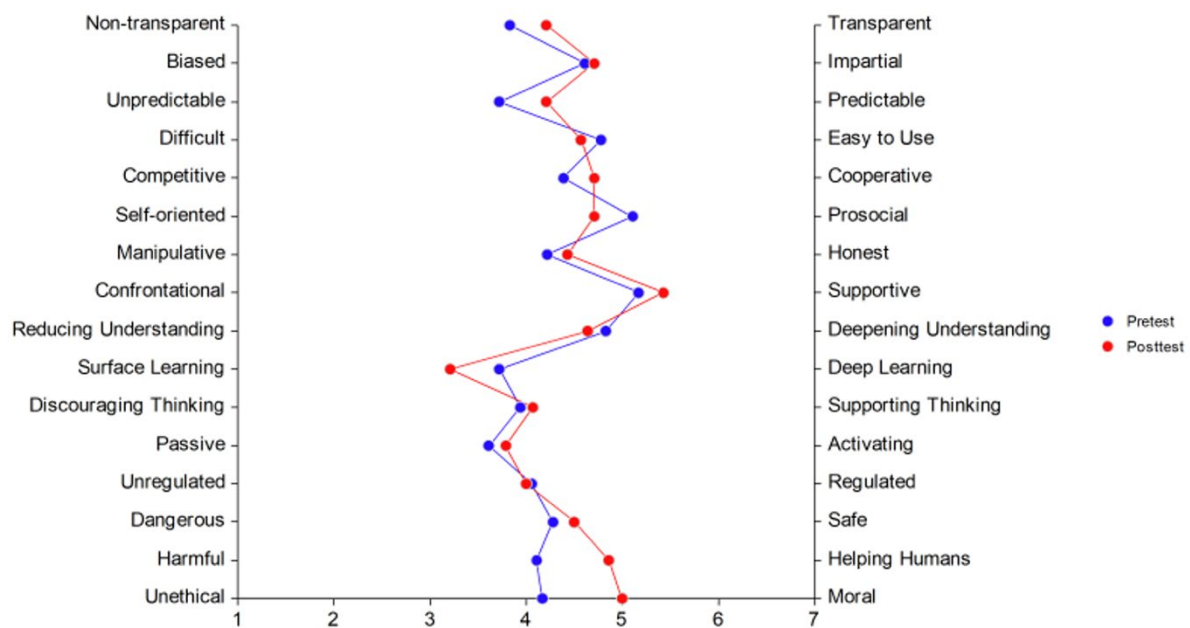


Figure 1: Semantic differential profile – comparison of pretest and posttest measurements

Visual comparison of the profiles suggests that the overall shape of the curves remained relatively similar across both measurements, indicating a certain degree of stability in students' perceptions of generative AI. Nevertheless, several partial changes can be observed.

Slight increases in mean values were identified for dimensions such as:

- Non-transparent – Transparent (3.83 → 4.21)
- Unpredictable – Predictable (3.72 → 4.21)
- Confrontational – Supportive (5.17 → 5.43)
- Dangerous – Safe (4.17 → 4.50)

In contrast, decreases were observed for dimensions such as:

- Self-oriented – Prosocial (5.11 → 4.71)
- Surface Learning – Deep Learning (3.72 → 3.21)

These differences represent descriptive changes in mean values only and do not indicate statistical significance by themselves. Therefore, statistical significance was subsequently examined using non-parametric methods.

3.3 Comparison of Pretest and Posttest Results Using the Mann–Whitney U Test

To examine differences between pretest and posttest measurements, the Mann–Whitney U test was employed. This procedure was selected because pretest and posttest responses could not be reliably matched at the individual participant level. Consequently, both measurements were analysed as two independent samples.

Furthermore, the data consisted of ordinal responses obtained through seven-point semantic differential scales, supporting the use of non-parametric statistical procedures.

The results of the Mann–Whitney U test did not reveal statistically significant differences between pretest and posttest measurements in any of the observed dimensions. In all cases, p-values exceeded the significance threshold of 0.05. The lowest p-value was observed for the *Unpredictable–Predictable* dimension ($p = 0.273$); however, even this result did not reach statistical significance. Effect sizes ranged from $r = 0.010$ to $r = 0.194$, corresponding to negligible to small effects.

Table 2: Mann–Whitney U test comparing Pretest and Posttest Measurements

Dimension	U	Z	p	r	Effect Interpretation
Non-transparent - Transparent	144	0.709	0.478	0.125	Small
Biased - Impartial	134.5	0.332	0.74	0.059	Negligible
Unpredictable - Predictable	154	1.096	0.273	0.194	Small
Difficult - Easy to Use	116.5	-0.373	0.709	0.066	Negligible
Competitive - Cooperative	139	0.509	0.611	0.09	Negligible
Self-oriented - Prosocial	105	-0.824	0.41	0.146	Small
Manipulative - Honest	122.5	-0.137	0.891	0.024	Negligible
Confrontational - Supportive	139.5	0.532	0.595	0.094	Negligible
Reducing Understanding - Deepening Understanding	115.5	-0.42	0.674	0.074	Negligible
Surface Learning - Deep Learning	106	-0.801	0.423	0.142	Small
Discouraging Thinking - Supporting Thinking	137	0.43	0.667	0.076	Negligible
Passive - Activating	143	0.666	0.506	0.118	Small
Unregulated - Regulated	124.5	-0.059	0.953	0.01	Negligible
Dangerous - Safe	146	0.801	0.423	0.142	Small
Harmful - Helping Humans	111.5	-0.564	0.573	0.1	Small
Unethical - Moral	143	0.666	0.505	0.118	Small

Note: $r = |Z|/\sqrt{N}$, where $N = 32$. Effect sizes were interpreted as negligible for $r < 0.10$, small for approximately $r \geq 0.10$, medium for $r \geq 0.30$, and large for $r \geq 0.50$.

No statistically significant differences were detected between the pretest and posttest measurements in any semantic differential dimension. Because the samples were small and unmatched, the minor differences in mean scores are exploratory descriptive patterns only and should not be interpreted as evidence of individual attitude change or an intervention effect. The results therefore provide no statistical basis for concluding that the intervention altered students' general attitudes towards generative AI.

4. Discussion

The aim of this study was to compare pretest and posttest perceptions of generative AI among pre-service mathematics teachers in the context of an educational intervention focused on AI principles, cognitive offloading, automation bias, and game theory. No statistically significant differences were detected in any semantic differential dimension. The observed mean differences were descriptive only and cannot be attributed to the intervention. Accordingly, this study provides no evidence that the short intervention altered students' general attitudes towards AI; at most, the descriptive patterns identify dimensions for further investigation. This interpretation corresponds with previous studies on AI literacy among future teachers, which suggest that educational interventions often influence awareness, critical reflection, and understanding of AI principles rather than immediately changing overall attitudes or trust in technology (Bilbao Eraña and Arroyo-Sagasta, 2025; Hsu, 2026).

One exploratory descriptive pattern was the lower posttest mean for the Surface Learning–Deep Learning dimension; this difference was not statistically significant. Although this pattern is compatible with greater awareness of risks related to cognitive offloading and excessive reliance on AI systems, the present design cannot determine its cause. The intervention also included explanations of the probabilistic nature of large language models and discussions of automation bias, but the data do not establish that these components produced more critical reflection. Related concerns have been reported in research on cognitive offloading, metacognitive calibration, and automation bias in AI-supported environments (Grinschgl et al, 2025; Zhai, Wibowo and Li, 2024; Lee et al, 2025; Nguyen, 2024; Parasuraman and Manzey, 2010).

From the perspective of mathematics teacher education, the findings suggest that AI literacy should not focus solely on technical use of tools but should also support the development of critical thinking and understanding of the limitations of generative systems. This is particularly important in mathematics and other exact disciplines, where correctness of results and logical consistency represent fundamental aspects of the educational process (Canonigo, 2024; Soygazi and Oğuz, 2023).

The findings should be interpreted in light of important design limitations. The study used a small convenience sample from a single programme, included no control group, and could not match individual pretest and posttest responses. The resulting independent-sample analysis had limited statistical power, particularly for detecting small effects, and the design does not support causal attribution. The small, non-random sample also limits generalisability beyond this cohort. The relatively short intervention may have further reduced the possibility of observing differences. In addition, the Prisoner's Dilemma functioned as a conceptual and pedagogical lens rather than an empirically operationalised model; the study therefore does not test game-theoretic predictions.

5. Conclusion

This study compared pretest and posttest perceptions of generative artificial intelligence among pre-service mathematics teachers in the context of an educational intervention focused on AI functioning and related cognitive risks.

No statistically significant differences were identified between the two measurements. The small differences in mean scores were descriptive only and do not demonstrate individual change or an effect of the intervention. Accordingly, the study cannot determine whether the intervention altered students' attitudes or understanding.

The study nevertheless illustrates how the Prisoner's Dilemma can be used as a conceptual pedagogical lens for discussing cognitive offloading and excessive reliance on AI in teacher education. It also highlights the relevance of connecting technical AI literacy with critical reflection, without claiming that the present intervention produced this outcome.

Future research should examine these findings using larger samples and longitudinal research designs.

Ethics Declaration

Participation in the study was voluntary and all participants were informed about the purpose of the study before providing informed consent.

AI Declaration

Generative AI tools (ChatGPT, OpenAI) were used for language refinement, text structuring, and stylistic editing. All scientific interpretation, analysis, and final decisions remained the responsibility of the authors.

References

- Bilbao Eraña, A. and Arroyo-Sagasta, A. (2025) "Fostering AI literacy in pre-service teachers: Impact of a training intervention on awareness, attitude and trust in AI", *Frontiers in Education*, Vol. 10, Article 1668078. DOI: 10.3389/feduc.2025.1668078.
- Canonigo, A.M. (2024) "Leveraging AI to Enhance Students' Conceptual Understanding and Confidence in Mathematics", *Journal of Computer Assisted Learning*, Vol. 40, No. 6, pp. 3215–3229. DOI: 10.1111/jcal.13065.
- Grinschgl, S. et al. (2025) "Cognitive offloading or cognitive overload? How AI alters the mental ecology of learning", *Frontiers in Psychology*, Vol. 13, Article 1699320. DOI: 10.3389/fpsyg.2025.1699320.
- Hsu, H.P. (2026) "Artificial Intelligence Literacy and Competency in Pre-Service Teacher Education", *Encyclopedia*, Vol. 6, No. 4, Article 76. DOI: 10.3390/encyclopedia6040076.
- Kasneci, E. et al. (2023) "ChatGPT for good? On opportunities and challenges of large language models for education", *Learning and Individual Differences*, Vol. 103, Article 102274. DOI: 10.1016/j.lindif.2023.102274.
- Lee, H. et al. (2025) "Learning Behaviors Mediate the Effect of AI-powered Support for Metacognitive Calibration on Learning Outcomes", in *Proceedings of the ACM Learning at Scale Conference*. DOI: 10.1145/3706598.3713960.
- Lin, Z. and Sohail, A. (2026) "Recalibrating academic expertise in the age of generative AI", *Patterns*, Vol. 7, No. 1, Article 101473. DOI: 10.1016/j.patter.2025.101473.
- Nguyen, T. (2024) "ChatGPT in Medical Education: A Precursor for Automation Bias?", *JMIR Medical Education*, Vol. 10, e50174. DOI: 10.2196/50174.
- Parasuraman, R. and Manzey, D.H. (2010) "Complacency and Bias in Human Use of Automation: An Attentional Integration", *Human Factors*, Vol. 52, No. 3, pp. 381–410. DOI: 10.1177/0018720810376055.
- Perc, M. et al. (2024) "Emergence of cooperation in the one-shot Prisoner's Dilemma when AI mediates interactions", *Journal of the Royal Society Interface*, Vol. 21, No. 218, Article 20240212. DOI: 10.1098/rsif.2024.0212.
- Soygazi, H. and Oğuz, A. (2023) "An Analysis of Large Language Models and LangChain in Mathematics Education", in *Proceedings of the 2023 International Conference on Educational Technology*. DOI: 10.1145/3633598.3633614.
- Zhai, C., Wibowo, S. and Li, L.D. (2024) "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review", *Smart Learning Environments*, Vol. 11, Article 28. DOI: 10.1186/s40561-024-00316-7.