Qwerty Journal

ONLINE FIRST

Beyond the Oracle: Toward a Critical Approach to the Use of ChatGPT in Education

Maria Ranieri [a], *Gabriele Biagini* [b]

Abstract

ChatGPT is often depicted either as an all-powerful tutor or as a threat to students' cognitive growth—both views ignoring the complex nature of learning. Under what conditions can teachers use ChatGPT critically and productively? We conducted an exploratory workshop with 147 Primary Education students over five weeks (10 hours total). Guided by revised Bloom's Taxonomy, each group completed iterative phases while interacting with ChatGPT 4 to collect information, refine texts, and design teaching scenarios. A mixed-methods design combined phase-by-phase satisfaction items with qualitative thematic analysis of metacognitive diaries (≈19,500 words) and group submissions. Qualitative evidence shows a progression from information retrieval and text refinement toward analysis and evaluation: students increasingly questioned AI outputs, triangulated sources, and justified accept/reject decisions. Quantitatively, self-reported satisfaction was high in early phases and dipped when tasks demanded higher-order reasoning, consistent with increased cognitive demands. Students learned to use ChatGPT as a collaborative tool supporting their own cognitive development rather than as an "oracle", demonstrating the importance of pedagogical scaffolding in AI-mediated learning.

Keywords: ChatGPT, Bloom's Taxonomy, Primary Education, Teacher Training, Critical Thinking

a University of Florence, Florence, Italy, https://orcid.org/0000-0002-8080-5436

b University of Florence, Florence, Italy, https://orcid.org/0000-0002-6203-122X

Corresponding author: gabriele.biagini@unifi.it

Introduction

Over the last decade, the rising interest in artificial intelligence (AI) has captivated all spheres of our societies (Singla et al., 2025; Stanford Institute for Human-Centered Artificial Intelligence, 2024), including higher education institutions (Ranieri & Gaggioli, 2025). Teachers and students are increasingly approaching AI tools with implications for educational methodologies and classroom dynamics in terms of both opportunities (e.g., personalization, intelligent tutoring or feedback automation) (Létourneau et al., 2025) and challenges (e.g., bias and fairness, surveillance concerns or transparency) (Idowu, 2025).

And yet, while ChatGPT use is increasing, critical literacy on AI is not growing at the same pace. On the one hand, teachers report a lack of training and adequate skills for the appropriate use of AI in teaching (UNESCO, 2024). On the other hand, students often show limited critical awareness when using tools such as ChatGPT and similar systems, with negative consequences at both the cognitive level (e.g., cognitive offloading) and the cultural level (e.g., reliance on unvalidated informational resources) (Gerlich, 2025). This uncritical approach—what we might term an "oracle approach"—manifests when students treat ChatGPT as an infallible source of knowledge, passively accepting its outputs through copy-paste behaviors without exercising critical judgment or verification (Cotton et al., 2024; Zhai et al., 2024).

In parallel, the current debate on AI for education is dominated by the tendency to treat AI as the leading factor influencing educational outcomes. This influence is often viewed through a dichotomic lens. On one end, AI is hailed as a transformative force leading to techno-solutionist approaches that can manifest in over-reliance on educational technologies to address challenges such as student engagement, personalized learning, or educational access without addressing deeper systemic issues such as curriculum quality, teacher training, or socio-economic disparities among students (Al-Zahrani, 2024; Chan & Lee, 2023).

On the other, sceptics of AI argue that the introduction of the artificial ingredient into the educational experience can lead to neglect of the human elements that are crucial in education, such as teacher-student relationships, motivation, and the development of critical thinking skills, potentially encouraging dependency among learners (Al-Zahrani, 2024; Selwyn, 2022).

Both viewpoints share a common underlying assumption: that AI, particularly ChatGPT, is the sole or primary agent in the educational process. However, this premise overlooks the multifaceted nature of education, which is influenced by a broad array of factors including, but not limited to, curricular content, pedagogical approaches, social interactions, institutional policies, and broader socio-economic contexts. Education is a complex, dynamic interplay of multiple elements where technology is just one part of the ecosystem.

To address these challenges, we cannot simply ban the use of ChatGPT among students. Assuming an educational perspective entails approaching the issues through pedagogical mediation. Both prohibition and the tendencies toward hype and oversimplification are far removed from the true aims of education. On the contrary, looking at AI through educational lens involves exercising the pedagogical mediation that is intrinsic to the teacher's role—a responsibility that cannot, and should not, be delegated to the machine. Pedagogical mediation involves multiple dimensions of the educational process. In this paper, we focus on the design and management of an educational intervention supported by ChatGPT, explicitly aimed at fostering students' critical and informed use of the tool. In particular, we designed a workshop based on Anderson and Krathwohl's revised Bloom's Taxonomy (Anderson & Krathwohl, 2001) and implemented at the University of Florence,

with the aim of analyzing how a scaffolding approach based on revised Bloom's taxonomy influences interaction patterns, cognitive processes, and pre-service teachers' perceptions in their use of ChatGPT for educational activities.

More specifically, to move beyond an "oracle approach" to the use of ChatGPT, we explored the following research questions:

RQ1: How does a revised Bloom's taxonomy-based scaffolding approach influence student teachers' interaction patterns and satisfaction with ChatGPT across different cognitive levels?

RQ2: What types of cognitive processes do student teachers demonstrate when using ChatGPT for educational tasks at progressively complex levels of the revised Bloom's taxonomy?

RQ3: How do student teachers perceive and evaluate ChatGPT's role in their learning process when structured through systematic pedagogical scaffolding?

Theoretical Background

There is a growing literature on the pedagogical affordances and educational uses of Generative AI, particularly language models like ChatGPT (Chocarro et al., 2021; Gao et al., 2022). In education, Limo et al. (2023) and Alam (2021) argue that ChatGPT enhances students' learning experiences by providing immediate, interactive, and tailored educational support. It serves multiple roles including as a research tool and a platform for generating educational content, according to Chen et al. (2023) and Abdelghani et al. (2022).

In higher education contexts, the applications expand significantly. At the university level, Barrett et al. (2019), Gill et al. (2024), and Kasneci et al. (2023) identify these models as instrumental in academic research and writing, helping to summarize texts and structure ideas. Bai et al. (2023) and UNESCO (2023) further suggest that these tools can refine research methodologies by proposing new research angles and compiling comprehensive resource lists. University students use AI tools for literature reviews, hypothesis generation, data analysis support, and academic writing across disciplines. The integration of AI in higher education raises important questions about assessment practices, academic integrity, and the development of critical scholarly skills.

For pre-service teacher education specifically, the implications are particularly significant. Future teachers must develop not only technical competence in using AI tools but also critical pedagogical judgment about when, how, and why to integrate such tools in their future classrooms (Guan et al., 2024; UNESCO, 2024). Teacher preparation programs face the dual challenge of preparing educators who can both leverage AI's pedagogical potential and recognize its limitations and risks. Pre-service teachers need to understand how AI tools might affect student learning processes, develop skills for designing AI-enhanced learning activities that maintain pedagogical integrity, and cultivate critical AI literacy that they can model for their future students (Pei et al., 2025; You & Zhang, 2025).

However, along with these pedagogical affordances, there are significant ethical and cognitive challenges to consider. Ethically, the use of AI in education raises concerns about biases in algorithms, data privacy, lack of transparency and many other ethical issues (Charisi et al., 2022; Floridi & Chiriatti, 2020; Holmes et al., 2022). Moreover, the cognitive implications of integrating AI tools like ChatGPT in educational settings cannot be underestimated. An important concern is the overreliance on AI, where learners may become overly dependent, potentially diminishing their motivation to learn and eroding their capacity to retain information. This dependency can lead to a deterioration in memory abilities due to the underutilization of these cognitive functions (Bai et al., 2023; Schemmer et al., 2022). Similarly, the convenience of accessing information through ChatGPT might impede the

development of critical thinking skills, essential for analytical, evaluative, and judgment processes in learning (Bai et al., 2023; Salim et al., 2023).

Combining Bloom's Taxonomy and Generative AI

Bloom's Taxonomy goes back to the Fifties (Bloom, 1956) and since then it has been a cornerstone in educational theory, providing a hierarchical classification of cognitive skills as a ground for a more systematic approach to design in education. In 2001, its revisions by Anderson and Krathwohl offered a more dynamic conception of educational objectives, distinguishing between types of knowledge and of cognitive processes. These latter are organised into six ascending levels, including Remember, Understand, Apply, Analyse, Evaluate, and Create, which guide educators in structuring learning activities that promote higher order thinking skills.

Beyond serving as a framework for classifying cognitive skills, Bloom's Taxonomy also functions as a design tool for defining learning objectives and structuring educational interventions. In this sense, it operates as a form of *pedagogical mediation*, enabling teachers to translate educational goals into meaningful and progressively complex learning experiences. Therefore, its integration with generative AI tools such as ChatGPT should not be seen as a purely technological enhancement but rather as the outcome of a *deliberate and pedagogically grounded intention* to use AI within a structured instructional design.

At the 'Remember' and 'Understand' levels, students use AI to gather information and clarify concepts, developing their foundational learning skills. In the 'Apply' phase, students utilize AI to help transform abstract knowledge into practical scenarios, enhancing their ability to implement theoretical concepts. At the 'Analyse' and 'Evaluate' stages, students engage with AI-generated content to examine diverse perspectives and problem-solving scenarios, developing their critical thinking and analytical skills.

While the 'Create' level represents the highest cognitive process in Bloom's revised taxonomy, this study focused primarily on the scaffold provided to develop critical evaluation skills (Remember through Evaluate) as a condition to move from a merely productive AI approach to a more creative one. The decision to emphasize critical analysis and evaluation over creation reflects both theoretical and practical reasons. While generation is the main feature of generative AI tools, critical thinking is less evident; yet generation without critical thinking does not amount to genuine creativity. Therefore, as a first step in our exploration, we sought to deepen our understanding of the cognitive dynamics stimulated by the interplay between pedagogical design and the affordances of generative AI. In addition, the time available for the workshop to adequately examine both critical and creative dimensions was not sufficient. Therefore, we decide to focus on creativity in a subsequent study. However, some creative elements were embedded throughout all phases, as students generated original texts, designed educational scenarios, and produced innovative solutions within each cognitive level.

Method

Research Design

This exploratory study employed a pre-experimental, one-group design with a mixed-methods approach, collecting data at each workshop phase. It aimed to understand how pedagogical scaffolding influence students' use of AI tools in terms of critical cognitive processes, rather than to establish causal relationships. This design choice was appropriate for this initial exploration of a novel pedagogical approach combining the revised Bloom's taxonomy with ChatGPT use.

Participants

The participants were 147 university students enrolled in the first year of Primary Education programs at the University of Florence during the 2023/2024 academic year. They were selected through convenience sampling to provide hands-on experience in integrating AI technology into educational settings. After removing 7 cases with incomplete data, the final sample comprised 140 participants: 135 females (96.4%) and 5 males (3.6%). The age distribution was: 134 participants aged 18-24 (95.7%) and 6 aged 25-34 (4.3%). Regarding educational background, 128 participants (91.4%) had completed high school, while 12 (8.6%) held a 3-year university degree. All participants provided informed consent for their data to be used for research purposes.

Procedure

The workshop was integrated into the mandatory course of Teaching Methods and Strategies, spanning five consecutive weeks with 2-hour weekly sessions (10 hours total).

Students worked collaboratively in small groups (3-5 participants, forming approximately 35-40 groups) on a comprehensive project that would unfold across four phases. The final expected output was the design of an educational scenario including: (1) a researched and refined conceptual foundation on a chosen pedagogical topic, (2) three alternative teaching scenarios generated and analysed for that topic, and (3) a critical evaluation comparing their self-assessment with ChatGPT's evaluation of their work. This scaffolded progression was designed to move students from information consumers to critical evaluators of both their own work and AI-generated content.

Students progressed through four phases aligned with Bloom's revised taxonomy:

Phase 1 - Remember (Week 1, 2 hours): Students selected educational topics of interest and used ChatGPT 4 to gather foundational information, producing structured texts. Popular topics included "Inclusive Teaching" (35 students, 25.0% of 140 participants), "Motivation in Learning Processes" (34 students, 24.3%), and "Personalization and Learning" (30 students, 21.4%).

Phase 2 - Understand (Week 2, 2 hours): Students refined their initial texts by identifying inconsistencies and gaps with ChatGPT's assistance, then conducted additional research beyond the AI tool to enhance their understanding and verify information.

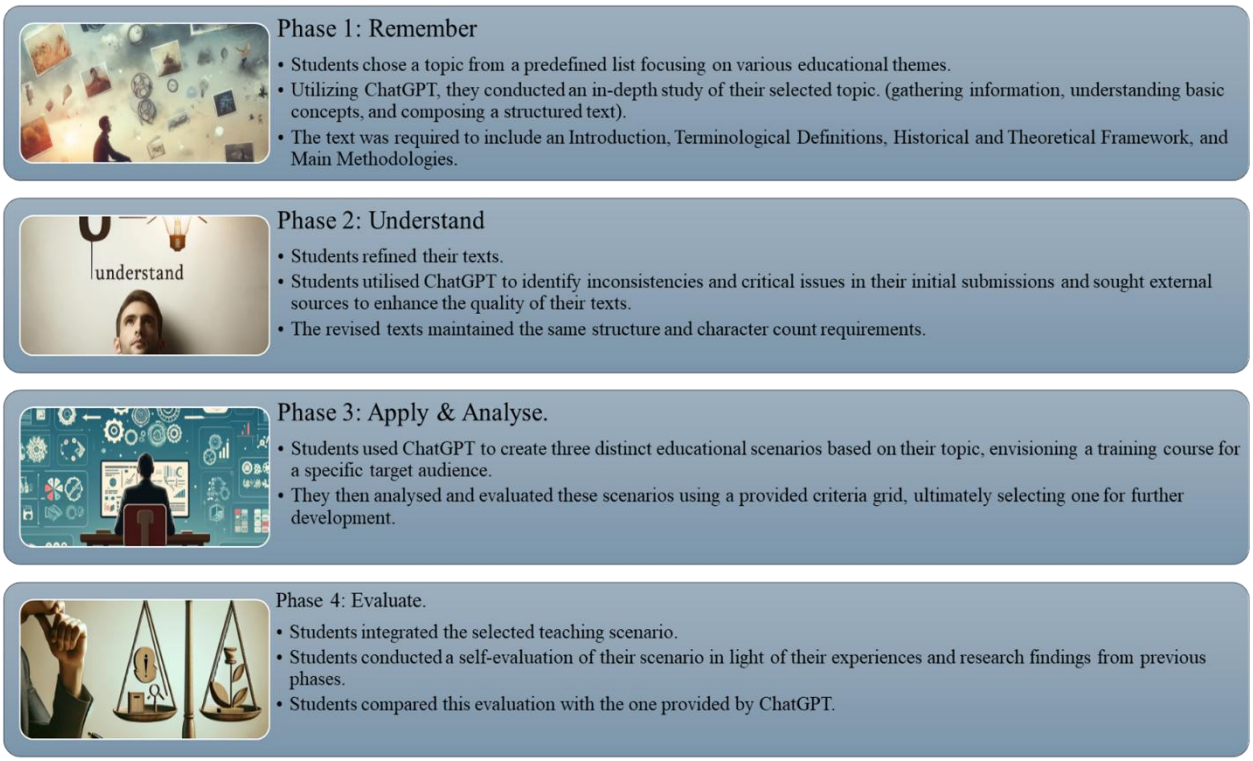
Phase 3 - Apply & Analyse (Weeks 3-4, 4 hours): Students used ChatGPT to generate three distinct educational scenarios based on their topics, then analysed each scenario against pedagogical criteria (age-appropriateness, feasibility, learning objectives alignment) to select the most suitable one.

Phase 4 - Evaluate (Week 5, 2 hours): Students critically compared their self-assessments with ChatGPT's evaluation of their work, reflecting on discrepancies and the overall learning process.

The next figure (Figure 1) holds the details of the four phases of the workshop.

Figure 1

Workshop phases details.



Materials and Instruments

All participants used ChatGPT 4 (OpenAI, accessed via web interface) with institutional accounts provided by the university, ensuring uniform access to the same AI capabilities.

With the aim of stimulating future teachers' reflection on the learning process and collect data on it, four structured diaries were created, one for each phase (i.e., Remember, Understand, Apply & Analyse, and Evaluate), including both closed-ended questions (satisfaction ratings on a 5-point Likert scale from 1=very unsatisfied to 5=very satisfied) and open-ended reflection prompts. Each diary focused on specific aspects of the learning process aligned with the corresponding revised Bloom's level. Participants were asked to complete the diary individually at the end of each phase and to upload it online through the platform. The diaries collected approximately 19,500 words of qualitative data. The university's LMS was used for workshop management, assignment submission, and diary collection, providing a centralized platform for all course activities. Figure 2 holds the details.

Figure 2

Metacognitive diaries Theme, Key questions, responses, word count.

Phase	*Theme*	*Key Questions (samples)*	*Nr. Responses*	*Question Count*	*Word Count*
1	Researching and organizing information	Open-ended: "What prompts did you use?", "Were you satisfied by the answer?", "Why?", "How did you modify the prompt?" Closed-ended: "Are you satisfied with the final output?"	140	8	~4,600
2	Improving text-based content	Open-ended: "How did you modify the prompt?", "List the alternative sources consulted", "What percentage of your text was revised based on suggestions from ChatGPT and alternative sources?" Closed-ended: "Are you satisfied with ChatGPT's suggestions?", "Are you satisfied with the final output?"	112	8	~5,200
3	Application and analysis of educational scenarios	Open-ended: "Briefly describe each of the educational scenarios you created with ChatGPT", "What was the main challenge in creating these scenarios?" Closed-ended: "Are you satisfied with the scenarios proposed by ChatGPT?", "Do you think the final text would have been better or worse without ChatGPT's support?"	112	8	~4,700
4	Evaluation of the learning process	Open-ended: "Are you satisfied with your final outcome? Explain why.", "How did you integrate the information and skills from the first three days?", "How does ChatGPT's evaluation differ from your self-evaluation?", "What changes would you apply to your learning process?", "If teaching someone to use ChatGPT for educational purposes, what advice would you give?"	100	8	~5,000

Data Analysis

Based on the data collected through the diaries, both quantitative and qualitative analyses were carried out. Descriptive statistics (frequencies and percentages) were calculated for closed-ended satisfaction questions included in the diary at the end of each phase. Thematic analysis was conducted inductively using Atlas.ti software (Version 23.2.1) on participants' reflections collected through the open-ended diary comments (=19,500 words) and Moodle submissions. A single coder systematically identified recurring themes and patterns following Mayring's (2014) qualitative content analysis approach. To ensure validity of the qualitative findings, inter-rater reliability was assessed on a random sample representing ≈10% of the total corpus (n=40 coded segments distributed across all four phases). A

second coder, experienced in qualitative research and educational technology, independently coded the sampled data using the established coding scheme. Agreement rates were calculated at two levels: macro-categories showed 88% agreement (Cohen's $\kappa = 0.82$), while thematic categories showed 82% agreement (Cohen's $\kappa = 0.79$). According to Landis and Koch (1977), these values indicate almost perfect and substantial agreement respectively, demonstrating acceptable inter-rater reliability. Disagreements were resolved through discussion between coders, and the resolved coding informed final interpretation. Quantitative and qualitative findings were triangulated to provide a comprehensive understanding of students' experiences across the four phases. All the results were reported in Table 1.

Results

Phase 1: Remember

In the initial phase, students employed ChatGPT to support their cognitive processes of information gathering and recall. The use of ChatGPT facilitated the initial gathering of information and construction of text for 121 participants (86.4% of 140 total participants). Figure 3 holds quotation details.

When asked about their appreciation for ChatGPT's responses, 97 participants provided detailed feedback (69.3% response rate from 140 total participants). Two macro-categories emerged from the thematic analysis:

Quality of Interaction encompassed responses from 59 participants (42.1% of 140 total participants), divided into three thematic areas:

- Exhaustiveness and precision (n=26, 18.6% of 140): Students noted that ChatGPT "elaborated in a complete manner" and provided "exhaustive and clear responses"
- Adaptability (n=15, 10.7% of 140): Participants appreciated how ChatGPT "demonstrated a significant understanding of our project's unique requirements, tailoring its responses accordingly"
- Dissatisfaction (n=18, 12.9% of 140): Some students felt responses were "too broad" or "generic"

Efficiency and Responsiveness included responses from 53 participants (37.9% of 140 total participants):

- Clarity and simplicity (n=15, 10.7% of 140): Students appreciated responses that were "clear, precise, and concise"
- Speed and efficiency (n=20, 14.3% of 140): Participants noted that "in very few seconds it illustrates all possible answers"
- Consistency and Relevance (n=18, 12.9% of 140): Students observed that "the text was well cohesive and respected the required structure".

Table 1.
Phase 1 details, thematic categories, sample quotation, responses, percentages of respondents.

Macro-categories	Definition	Thematic Categories	Sample Quotation	NR of Ref	% of Respondents
Quality of Interaction	Refers to the depth and comprehensiveness of engagement facilitated by the system. This macro-category encompasses the clarity, coherence, and relevance of the responses, highlighting the system's ability to provide exhaustive and precise information tailored to user inquiries.	Exhaustiveness and precision	"Elaborated in a complete manner" and "exhaustive and clear in the given response."	26	26.80%
		Low quality	"It was a too broad response", "The definition given by ChatGPT was too generic and formulated with a non-specific lexicon	18	18.56%
		Adaptability	"… demonstrated a significant understanding of our project's unique requirements, tailoring its responses accordingly."	15	15.46%
Efficiency and Responsiveness	Focuses on the system's ability to provide quick and efficient responses, emphasizing the speed at which it operates and the efficiency in delivering coherent	Speed and efficiency.	"In very few seconds it illustrates all possible answers" and "coherence of responses received in a	20	20.62%
		Consistency and relevance		18	18.56%
		Clarity and simplicity		15	15.46%

	and relevant information.	very short time." "The text was well cohesive and respected the required structure" and "the generated response was in line with our requests." "Clear, precise, and concise" and “highlighted the exhaustiveness and clarity of the information provided”.

Phase 2: Understand

In the second phase, students engaged in deeper cognitive processing, using ChatGPT to identify inconsistencies and refine their understanding. When asked about satisfaction with ChatGPT's suggestions and modifications, 112 participants responded positively. Figure 4 holds quotation details.

From the 82 participants who provided detailed feedback (73,2% of 112), three macro-categories emerged:

Content Enhancement and Expansion (37 participants, 33,0% of 112):

- Text Enhancement (37 participants, 33,0% of 112): Students noted ChatGPT "managed to improve the lexicon of the text, structuring it in a more coherent way"
- Amplification and Examples (n = 15, 13,4% of 112): Participants observed it "expanded and corrected our speeches making them more readable and including examples"

Creativity and Innovation (51 participants, 45,5% of 112):

- Better Text Quality (n = 25, 22,3%)
- Ideas and Creativity (n = 20, 17,9%)
- Enhanced Creativity and Originality (n = 6, 5,4%)

Analytical and Critical Enhancement (13 participants, 11,6% of 112):

- Critical Analysis and Text Improvement: Students recognized ChatGPT's role in "identifying and addressing content gaps and inconsistencies".

Table 2.
Phase 2 details, thematic categories, sample quotation, responses, percentages of respondents.

Macro-categories	Definition	Thematic Categories	Sample Quotation	NR of Ref	% of Respondents
Content Enhancement and Expansion	Encompasses improvements in text quality, including coherence, lexicon, and the inclusion of relevant examples. It highlights the system's capacity to enrich the text, making it more informative and engaging for the reader.	Better text quality	"Thanks to ChatGPT the text is more fluid and understandable." Another noted the potential drawbacks without ChatGPT, mentioning it would have been "less fluid, would have had less information and would have been elaborated in more time."	25	45.45%
		Text enhancement	"Managed to improve the lexicon of the text, structuring it in a more coherent way," and "rework the text adding useful information."	22	26.83%
		Amplification and examples	"Because it expanded and corrected our speeches making them more readable and including examples," and "Because it expanded the answers it had given us making them more complete using examples."	15	18.29%
Creativity and Innovation	Focuses on the system's ability to introduce new, creative ideas and suggestions that contribute to the text's improvement, emphasizing originality and inventive solutions to enhance the content.	Ideas and creativity	"Because it gave positive ideas to improve the text interestingly and correctly," and "We managed to improve our text."	20	24.39%
		Enhanced Creativity and Originality	"Because we would have had more information from different sources and the text would have come out more original," suggesting that ChatGPT's diverse information sources	6	10.91%

Macro-categories	Definition	Thematic Categories	Sample Quotation	NR of Ref	% of Respondents
			contributed to a more unique and creative output.		
Analytical and Critical Enhancement	Highlights the system's capability to perform critical analysis, identify inconsistencies, and suggest improvements. It emphasizes a deep understanding of the content, providing insightful feedback to refine and advance the text's quality.	Critical analysis and text improvement	"Because it found our inconsistencies," and "It managed to identify the points not thoroughly explored and gave us additional information without ever being repetitive."	13	15.85%
Information Accessibility and Structure	Addresses the system's role in making information more accessible and improving the structural coherence of the text. It encompasses the ability to gather and consolidate diverse information sources, thereby facilitating a more structured and cohesive content creation process.	Accessibility of information	"Because many pieces of information without ChatGPT we would not have found." Another highlighted the difficulty in consolidating sources, saying, "Because it would have been more difficult to find all the sources in a single document."	17	30.91%
		Structure improvement	"Because it was useful to confer structure and cohesion to the text." Another expressed reliance on ChatGPT for starting the writing process: "Without ChatGPT I wouldn't even know where to start."	7	12.73%

Phase 3: Analyse and Apply

In this phase, students demonstrated analytical skills through scenario creation and evaluation. Student satisfaction with ChatGPT's scenario generation varied: 98 participants responded to the satisfaction question (87.5% response rate from 112 participants). Among these respondents: 45 students (40.2% of 112) were satisfied with all three scenarios proposed, 31 students (27.7% of 112) were satisfied with two scenarios, 21 students (18.8% of 112) were satisfied with at least one scenario, and one student (0.9% of 112) was not satisfied with any scenario. The non-response from 14 participants (12.5%) in this phase represents a notable drop in participation that warrants discussion in the limitations section. Figure 5 holds quotation details.

From the 74 participants who provided qualitative feedback (66.1% of 112), two macro-categories emerged:

Scenario Design and Implementation (42 participants, 37.5% of 112):

- Realism (n=23, 20.5% of 112): The system's capacity to generate scenarios "both engaging for students and feasible for classroom implementation"
- Detailed scenarios (n=19, 17.0% of 112): ChatGPT created "interesting and plausible scenarios"

Educational Creativity (32 participants, 28.6% of 112):

- Coherence with educational themes (n=18, 16.1% of 112): Scenarios "remained aligned with the educational content"
- Creativity (n=14, 12.5% of 112): Scenarios were "not too complicated for children attending primary school"

Table 3.
Phase 3 details, , thematic categories, sample quotation, responses, percentages of respondents.

Macro-categories	Definition	Thematic Categories	Sample Quotation	NR of Ref	% of Respondents
Scenario Design and Implementation	This macro-category encompasses the creation and application of educational scenarios, emphasizing realism, detail, and alignment with educational themes. It highlights the system's ability to generate content that is both engaging and feasible within an educational setting.	Realism	"Because ChatGPT was able to create realistic and effective educational scenarios," and noting their feasibility in classroom settings, "They are all feasible in class depending on the timing available and at the discretion of the teacher."	23	31.08%
		Detailed scenarios	"Because based on the targets we provided, it created three interesting and plausible scenarios." However, there was also feedback on the need for specificity, as another pointed out, "Only one of the scenarios was too generic and abstract."	19	25.68%
Educational Creativity	Focuses on the inventive aspect of educational content creation, particularly the development of scenarios that are not only creative but also age-appropriate and engaging for the intended audience, such as primary school children.	Coherence with educational themes	"Because they are coherent with the topic discussed," and "All scenarios were coherent with the topic discussed."	18	24.32%
		Creativity	"Because the scenarios proposed are creative and suitable for the age of the children.", "All the scenarios seemed	14	18.92%

Macro-categories	Definition	Thematic Categories	Sample Quotation	NR of Ref	% of Respondents
			not too complicated for children attending primary school."		

Phase 4: Evaluate

In the final phase, students demonstrated their ability to critically evaluate both their own work and ChatGPT's assessments. From 75 participants who provided feedback (75.0% response rate from 100), three macro-categories emerged:

Educational Scenario Development (48 participants, 48.0% of 100):

- Completeness (n=18, 18.0% of 100)
- Applicability of educational scenarios (n=14, 14.0% of 100)
- Alignment with Personal Evaluation (n=16, 16.0% of 100)

Satisfaction and Critical Evaluation (40 participants, 40.0% of 100):

- Final output satisfaction (n=14, 14.0% of 100): “Thanks to ChatGPT, we managed to create a good piece of work in a short time”
- Satisfaction with ChatGPT’s Evaluation (n=14, 14.0% of 100)
- Critical Perspective on ChatGPT’s Evaluation (n=12, 12.0% of 100)

Evaluation and Methodological Approach (55 participants, 55.0% of 100):

- Differences in Specific Aspects of Evaluation (n=4, 4.0% of 100)
- Lack of Influence on Personal Evaluation (n=8, 8.0% of 100)
- Formulating Clear and Specific Prompts (n=22, 22.0% of 100)
- Integration with Personal Insights (n=21, 21.0% of 100)

Table 4.
Phase 4 details, thematic categories, sample quotation, responses, percentages of respondents.

Macro-categories	Definition	Thematic Categories	Sample Quotation	NR of Ref	% of Respondents
Educational Scenario Development	Encompasses the creation, applicability, and alignment of educational scenarios with educational goals and personal evaluations. It highlights ChatGPT's ability to craft realistic and effective scenarios suited for classroom use.	Completeness Alignment with Personal Evaluation Applicability of educational scenarios	"Yes, because it was complete in most of the requested topics." "ChatGPT's evaluation differs only slightly from my self-assessment..." "Yes, because ChatGPT has managed to create realistic and effective educational scenarios."	18 16 14	39.13% 34.78% 30.43%
Critical Evaluation	Focuses on the satisfaction with the final outputs and ChatGPT's evaluations, including a critical perspective on these evaluations. It underscores the importance of assessing the quality and impact of the educational content.	Final output satisfaction Satisfaction with ChatGPT's Evaluation Critical Perspective on ChatGPT's Evaluation	"Yes, I am very satisfied because thanks to ChatGPT, we managed to create a good piece of work in a short time." "Yes, because Chat GPT has evaluated our scenario with satisfactory scores." "We have had a more critical view than ChatGPT regarding the development objectives of empathy..."	14 14 12	30.43% 30.43% 26.09%

Evaluation and Methodological Approach	Covers the approach to evaluating educational scenarios, including differences in evaluation, lack of influence on personal evaluation, and the importance of formulating clear prompts and integrating personal insights.	Formulating Clear and Specific Prompts	"I would advise formulating the question well to the AI..."	22	29.33%
		Integration with Personal Insights	"To use ChatGPT always integrating something of your own."	21	28.00%
		Lack of Influence on Personal Evaluation	"Not at all, because we did the personal evaluation first..."	8	16.36%
		Differences in Specific Aspects of Evaluation	"Differs drastically only in the field of technological supports..."	4	8.70%
Critical Engagement and Improvement	Highlights the critical interaction with ChatGPT for enhancing educational scenarios and the positive influence of ChatGPT's guidance, along with the validation of evaluations and the verification of information.	Verification and Critical Analysis	"To verify in detail the accuracy and relevance of the information suggested by the chat."	24	32.00%
		Approaching ChatGPT with Caution	"I would advise not to rely too much on the answers proposed by it and to do a careful analysis of the structure and form."	8	16.36%
		Critical Interaction with ChatGPT		6	13.04%
		Positive Influence and Guidance	"I interacted with ChatGPT critically, reworking its answers in the right way."	4	8.18%
		Evaluation Validation and Confirmation	"It influenced a lot since I had never created it before, its support was a source of guidance for me."	4	8.18%
			"My self-evaluation coincided largely		

		with that of ChatGPT."		

Table 5.
Satisfaction responses by workshop phase with descriptive statistics. Satisfaction was measured on a 5-point Likert scale (1=very unsatisfied, 5=very satisfied). Phase 3 used a categorical response format for scenario evaluation.

Workshop Phase	Response Categories	N (%)	Mean	SD
Phase 1 (Remember) n=140	Satisfied (3-4-5 on Likert)	122 (87.1%)	4.02	0.54
	Not satisfied (1-2 on Likert)	18 (12.9%)		
Phase 2 (Understand) n=112	Satisfied (3-4-5 on Likert)	97 (86.6%)	4.16	0.55
	Not satisfied (1-2 on Likert)	15 (13.4%)		
Phase 3 (Apply & Analyse) n=112	All three scenarios	54 (48.2%)	3.61	0.83
	Two scenarios	23 (20.5%)		
	One scenario	33 (29.5%)		
	No scenario	2 (1.8%)		
Phase 4 (Evaluate) n=100	Satisfied (3-4-5 on Likert)	100 (100%)	4.10	0.64

Discussion

This exploratory study investigated whether a Bloom's taxonomy-based scaffolding approach could help student teachers move beyond passive "oracle" reliance on ChatGPT toward critical engagement. Three key findings emerged. First, interaction patterns evolved significantly: students progressed from appreciating ChatGPT primarily for speed and comprehensiveness in early phases (Phase 1: M=4.02, SD=0.54; Phase 2: M=4.16, SD=0.55) to actively evaluating AI outputs against pedagogical criteria in later phases, where satisfaction declined notably (Phase 3: M=3.61, SD=0.83) before partially recovering (Phase 4: M=4.10, SD=0.64). Second, qualitative analysis revealed a developmental trajectory in cognitive processes, from basic information retrieval through content refinement to metacognitive reflection and critical evaluation. Third, students' perceptions shifted fundamentally from viewing ChatGPT as an authoritative information provider to recognizing it as a tool supporting—but not replacing—their own cognitive processes and professional judgment. These findings demonstrate that with appropriate pedagogical scaffolding, the "oracle approach" can be transformed into critical partnership.

Regarding how the Bloom's-based scaffolding influenced interaction patterns and satisfaction (RQ1), our findings reveal that the structured progression successfully shaped students' engagement with

ChatGPT while also exposing the cognitive demands of higher-order thinking. The consistently high satisfaction in initial phases (Phase 1: M=4.02; Phase 2 reaching peak: M=4.16) suggests students found the scaffolding supportive for foundational tasks. However, the more nuanced satisfaction patterns in later phases—particularly in Apply/Analyse where satisfaction dropped significantly to M=3.61 (a decline of 0.55 points from Phase 2) and only 48.2% were satisfied with all three AI-generated scenarios (with 29.5% satisfied with only one scenario)—indicates that students were genuinely exercising independent judgment rather than uncritically accepting all outputs. Notably, the increased response variability in Phase 3 (SD=0.83 compared to ~0.55 in other phases) further demonstrates the diverse cognitive experiences students encountered when engaging with complex analytical tasks.

This progression directly addresses the concerns raised by Cotton et al. (2024) and Zhai et al. (2024) regarding students' uncritical acceptance of AI outputs through passive copy-paste behaviors. While students initially exhibited characteristics typical of what we termed "oracle-like dependence"—with 87.1% satisfied in Phase 1 and 86.6% in Phase 2, appreciating ChatGPT primarily for its efficiency and comprehensiveness—the scaffolded approach fostered increasingly critical engagement. These results align with Schemmer et al.'s (2022) framework of "appropriate reliance" in human-AI decision-making, extending it to educational contexts. Our data suggest that appropriate reliance develops progressively rather than emerging spontaneously: students required structured guidance through increasingly complex cognitive tasks to develop the capacity to question, triangulate, and critically evaluate AI outputs.

Regarding RQ2, the qualitative data revealed that students did not simply perform tasks at different cognitive levels; rather, they internalized increasingly sophisticated approaches to AI interaction. The quantitative pattern supports this developmental trajectory: high initial satisfaction (Phases 1-2: M>4.00), a significant dip during complex analytical work (Phase 3: M=3.61), and partial recovery during meta-evaluation (Phase 4: M=4.10). By the final phase, where 100% of students reported satisfaction despite the cognitive demands of evaluating both their own work and the AI's assessments, students engaged in meta-evaluation, critically assessing not only their own work but also the AI's evaluation itself—a form of higher-order thinking that represents genuine critical AI literacy, extending previous research on ChatGPT's pedagogical affordances (Kasneci et al., 2023; Gill et al., 2024; Barrett et al., 2019) by demonstrating not merely that AI can support various cognitive levels, but that structured progression through these levels shapes students' capacity for critical engagement. The finding addresses cognitive concerns raised by Bai et al. (2023) and Salim et al. (2023) regarding cognitive offloading and diminished critical thinking. Rather than undermining cognitive development, our scaffolded approach appeared to support it by requiring active engagement with AI outputs at each level, preventing the passive dependency that Gerlich (2025) warns can erode critical thinking capacities.

Students' evolving perceptions of ChatGPT's role represent perhaps the most significant finding regarding critical AI literacy development (RQ3). The shift from viewing ChatGPT as an authoritative information source to recognizing it as a collaborative tool supporting their own cognitive processes mirrors exactly the kind of pedagogical judgment that Guan et al. (2024), Pei et al. (2025), and You and Zhang (2025) identify as essential for future teachers. The satisfaction pattern across phases (peak at M=4.16, dip to M=3.61, recovery to M=4.10) reflects not frustration with the tool but rather students' growing awareness of its appropriate use within pedagogical contexts. Our participants did not simply learn to use an AI tool; they developed critical literacy about AI's capabilities and limitations within educational contexts. This perception shift has particular significance for pre-

service teacher education. The emphasis on evaluating AI-generated scenarios against pedagogical criteria (age-appropriateness, feasibility, learning objectives alignment) required students to position themselves as educational professionals whose judgment supersedes technological outputs.

Our results underscore a fundamental principle often overlooked in discussions of educational technology: the quality of learning with AI depends less on the tool's capabilities than on the pedagogical framework mediating its use. The revised Bloom's taxonomy functioned not merely as a task sequence but as a form of pedagogical mediation that shaped students' evolving relationship with the AI tool. The quantitative evidence—particularly the satisfaction decline in Phase 3 coupled with increased variability (SD=0.83)—demonstrates that cognitive scaffolding based on Bloom's taxonomy successfully challenged students rather than simply facilitating task completion. These findings address both extremes of the current debate—the view of AI as either educational panacea or existential threat (Al-Zahrani, 2024). Our evidence suggests that AI's educational impact is neither inherently positive nor negative but rather contingent on the pedagogical intentions and instructional designs that guide its integration.

The collaborative group structure and metacognitive diary requirements likely contributed significantly to this development. Collective discussion of AI outputs provided opportunities for students to articulate, challenge, and refine their thinking—forms of social learning that remain essential even in AI-mediated education. The structured reflection activities encouraged metacognitive awareness, helping students recognize their own agency in the learning process rather than attributing outcomes primarily to the AI tool. These findings suggest that developing critical AI literacy requires not just structured tasks but also deliberate opportunities for reflection and dialogue about the learning process itself, aligning with recommendations from UNESCO (2023) regarding AI integration in higher education.

Furthermore, our emergent finding that creative elements appeared throughout all phases, not only at a hypothetical final level, challenges the strictly hierarchical assumptions of Bloom's taxonomy in the context of AI-mediated learning. As suggested by the reviewer, this invites reflection on alternative theoretical models. We propose that a more flexible, perhaps cyclical or "networked" model, may be more appropriate. In such a model, students iterate constantly between different cognitive levels (e.g., Creating a prompt, Evaluating the AI output, and Understanding new information) rather than proceeding linearly. This study's findings support the need for theoretical models that acknowledge the simultaneous occurrence of higher- and lower-order processes during generative AI interaction.

Limitations

This exploratory study has several important limitations. The pre-experimental design without a control group fundamentally limits our ability to establish causal relationships—we cannot definitively attribute observed changes to the Bloom-scaffolded approach versus other factors. Our reliance on self-reported satisfaction rather than objective performance measures limits understanding of actual learning outcomes. The convenience sample presents multiple limitations: single Italian university context, and homogeneity in age and educational background all restrict generalizability. The qualitative analysis was strengthened through inter-rater reliability assessment, while this validation supports the trustworthiness of our thematic analysis, readers should note that the remaining 90% of data was coded by a single researcher following established protocols. Future research would benefit from complete dual coding of all transcripts to further enhance validity, though the strong agreement rates on our sample provide confidence in the coding approach. Varying response rates across phases, particularly 30% non-response in Phase 3 and continued attrition in

Phase 4, mean that non-response bias cannot be ruled out; it is possible that students who did not complete the diaries had somewhat different experiences from those who did, and results should therefore be interpreted with caution. Several factors may have contributed to this attrition. First, the workshop's progressive increase in cognitive demands may have led some students to disengage, particularly those who found higher-order tasks more challenging. Second, the time investment required for completing metacognitive diaries (particularly in later phases) may have exceeded some students' available time, especially given concurrent coursework demands. Third, technical issues with the LMS platform were occasionally reported, potentially preventing some submissions. Our focus on four rather than six Bloom's levels represents incomplete application of the framework. The emergent finding that creative elements appeared throughout all phases challenges strictly hierarchical assumptions and suggests need for more flexible theoretical models acknowledging simultaneous occurrence of higher- and lower-order processes. The inability to analyze ChatGPT interaction logs due to technical constraints represents a missed opportunity for objective behavioral data.

Conclusions

Despite these limitations, this exploratory study offers several practical implications for educators integrating AI tools in teacher education. The structured progression through Bloom's levels helped students develop from passive information consumers to critical evaluators, suggesting that AI tools require thoughtful pedagogical frameworks rather than mere technical access.

The collaborative group structure allowed students to discuss and challenge AI outputs collectively, suggesting that social learning remains important in AI-mediated education and that collective critical analysis may support development of individual critical dispositions. The 10-hour workshop format indicates that developing critical AI literacy requires substantial time investment—not quick training sessions—and that this investment yields progressive development of sophisticated critical capacities. The metacognitive diaries encouraged students to reflect on their AI use, suggesting that structured reflection activities should accompany AI integration to support development of metacognitive awareness essential for critical AI partnership.

For teacher education programs considering AI integration, this study suggests that success requires more than simply providing access to AI tools—it demands thoughtful pedagogical design, substantial time investment, opportunities for collaborative learning and reflection, and explicit focus on critical evaluation skills. As AI becomes increasingly prevalent in education, developing future teachers' capacity to use these tools critically—rather than as "oracles"—becomes essential. The title of our study encapsulates our key finding: when properly scaffolded, students can move beyond viewing ChatGPT as an all-knowing authority to understanding it as a tool that supports, but does not replace, their own cognitive processes and professional judgment. This shift from dependency to critical partnership represents the kind of AI literacy that future educators will need to navigate an increasingly AI-enhanced educational landscape while maintaining pedagogical integrity and fostering genuine learning.

References

Abdelghani, R., Wang, Y.-H., Yuan, X., Wang, T., Sauzéon, H., & Oudeyer, P.-Y. (2022). *GPT-3-driven pedagogical agents for training children's curious question-asking skills.* arXiv. https://arxiv.org/abs/2211.14228

Alam, A. (2021). Should robots replace teachers? Mobilisation of AI and learning analytics in education. In B. Patil, J. Save, & S. Deshmukh (Eds.), *Proceedings of the International Conference on Advances in Computing, Communication, and Control* (pp. 1–12). IEEE. https://doi.org/10.1109/ICAC353642.2021.9697300

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. Heliyon, 10(9), e30696. https://doi.org/10.1016/j.heliyon.2024.e30696

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Allyn & Bacon.

Bai, L., Liu, X., & Su, J. (2023). ChatGPT: The cognitive effects on learning and memory. *Brain-X, 1*, e30. https://doi.org/10.1002/brx2.30

Barrett, M., Branson, L., Carter, S., DeLeon, F., Ellis, J., Gundlach, C., & Lee, D. (2019). Using artificial intelligence to enhance educational opportunities and student services in higher education. *Inquiry: The Journal of the Virginia Community Colleges, 22*(1), 11–19.

Bloom, B. S. (1956). *Taxonomy of educational objectives: Handbook I, The cognitive domain*. David McKay.

Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learning Environments, 10*(1), 60. https://doi.org/10.1186/s40561-023-00269-3

Charisi, V., Chaudron, S., Di Gioia, R., Vuorikari, R., Escobar-Planas, M., Sanchez, I., & Gomez, E. (2022). *Artificial intelligence and the rights of the child: Towards an integrated agenda for research and policy* (EUR 31048 EN). Publications Office of the European Union. https://doi.org/10.2760/012329

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers, 25*, 161–182. https://doi.org/10.1007/s10796-022-10291-4

Chocarro, R., Cortinas, M., & Marcos-Matas, G. (2021). Teachers' attitudes towards chatbots in education: A technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies*, 1–19. https://doi.org/10.1080/03055698.2021.2010880

Cotton, D. R., Cotton, P. A., & Shipway, J. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv*. https://doi.org/10.1101/2022.12.23.521610

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Gill, S.S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ... (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems*, *4*, 19-23. https://doi.org/10.1016/j.iotcps.2023.06.002

Guan, L., Zhang, Y., & Gu, M. M. (2024). Pre-service teachers preparedness for AI-integrated education: An investigation from perceptions, capabilities, and teachers' identity changes. *Computers and Education: Artificial Intelligence, 8*, 100341. https://doi.org/10.1016/j.caeai.2024.100341

Holmes, W., Persson, J., Chounta, I. A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Council of Europe.

Idowu, A. (2025). Algorithmic bias in educational systems: Examining the impact of AI-driven decision. *World Journal of Advanced Research and Reviews, 25*(1), 2012-2017. https://doi.org/10.30574/wjarr.2025.25.1.0253

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *NPJ Science of Learning*, *10*(29). https://doi.org/10.1038/s41539-025-00320-7

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

Limo, F., Tiza, D. R. H., Roque, M., Mariki, K., & Liana, J. J. (2023). Personalized tutoring: ChatGPT as a virtual tutor for personalized learning experiences. *Social Space, 23*(1), 293–312.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. Self-published. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173

Pei, B.; Lu, J.; Jing, X. (2025) Empowering preservice teachers' AI literacy: Current understanding, influential factors, and strategies for improvement. Comput. Educ. Artif. Intell. 2025, 8, https://doi.org/10.1016/j.caeai.2025.100406

Ranieri, M., & Gaggioli, C. (2025). *Innovazione didattica e ambienti inclusivi all'università: Dalle competenze digitali all'intelligenza artificiale*. Edizioni ETS.

Salim, A., Jr., Allen, M., Mariki, K., Masoy, K. J., & Liana, J. J. (2023). Understanding how the use of AI decision support tools affect critical thinking and over-reliance on technology by drug dispensers in Tanzania. *arXiv*. https://doi.org/10.48550/arXiv.2302.09487

Schemmer, M., Hemmer, P., Kühl, N., Benz, C., & Satzger, G. J. (2022). Should I follow AI-based advice? Measuring appropriate reliance in human–AI decision-making. *arXiv*. https://doi.org/10.48550/arXiv.2204.06916

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education, 57*(4), 620–631. https://doi.org/10.1111/ejed.12501

Singla, A., Sukharevsky, A., Yee, L., & Chui, M. (2025, March). *The state of AI: How organizations are rewiring to capture value*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Stanford Institute for Human-Centered Artificial Intelligence. (2024). *The AI Index Report 2024*. Stanford University. https://hai.stanford.edu/ai-index/2024-ai-index-report

UNESCO. (2023). *Harnessing the era of artificial intelligence in higher education: A primer for higher education stakeholders*. https://unesdoc.unesco.org/ark:/48223/pf0000382740

UNESCO. (2024). *AI competency framework for teachers*. UNESCO. https://doi.org/10.54675/ZJTE2084

You, Y., & Zhang, Y. (2025). Developing AI literacies and negotiating professional identities: A study of pre-service English teachers in a ChatGPT-facilitated pedagogy. *Computer Assisted Language Learning*. https://doi.org/10.1515/jccall-2025-0007

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(28). https://doi.org/10.1186/s40561-024-00316-7