

University Students' Use of ChatGPT in Critical Thinking Tasks: Prompts, Performance, and Ethical Implications of Human-AI Collaboration in Higher Education

DARIA PIPA¹, MI CHAN HTAW², NAMKANG SRIWATTANAROTHAI³, CHAILERD PICHITPORNCHAI^{4*}^{1,2,3}Institute for Innovative Learning, Mahidol University, Nakhon Pathom, Thailand⁴Institute for Innovative Learning, Mahidol University, Nakhon Pathom, Thailand, and Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, ThailandEmail: ¹dariapi14@gmail.com, ²michanhtaw.mct@gmail.com, ³namkang.sri@mahidol.ac.thOrchid Id number: ¹<https://orcid.org/0009-0004-0152-1104>, ²<https://orcid.org/0009-0005-5803-3846>, ³<https://orcid.org/0000-0003-3082-0782>,⁴<https://orcid.org/0000-0003-2574-3858>Corresponding Author*: chailerd.pic@mahidol.ac.th**ABSTRACT:**

As universities increasingly integrate generative AI, its impact on foundational skills such as critical thinking remains contested. This study explored how thirty 18–23-year-old undergraduates at a central Thai university used ChatGPT-4o while completing critical thinking tasks involving analysis, evaluation, inference, decision-making, and metacognition. The primary objective was to determine in which task categories students relied on AI the most. Results showed that students most often turned to AI for higher-order processes, particularly inference, decision-making, and metacognitive reflective tasks. The study also examined how prompting strategies and the degree of AI use influenced performance. Four prompting strategies were identified, with filtering and paraphrasing of AI outputs linked to stronger outcomes, especially in metacognition, while direct copy-paste correlated with weaker performance. Lastly, the proportion of students who accurately estimated their AI contribution to task output was estimated. Results showed generally accurate self-assessment, though cases of over- and underestimation raised ethical concerns about transparency and academic honesty in human-AI collaboration. Overall, the findings offer empirical insights into students' use of AI in critical thinking, highlighting both opportunities and risks for transformative learning. They highlight the need for pedagogical strategies and ethical guidelines that foster authentic reasoning while positioning AI as a supportive tool for sustainable higher education.

KEYWORDS: critical thinking, writing skills, educational needs, generative AI, higher education.

1) Introduction:

Critical thinking skills are important in 21st century learning for students, helping them solve complex issues that arise in our multicultural and interconnected global society [1]. Examining how students interact with generative artificial intelligence (GenAI) technologies is crucial because it reveals not only their learning processes but also how they think and produce academic work. This is especially important as GenAI becomes increasingly integrated into higher education. Critical thinking abilities are necessary so that students are able to assess AI-generated content critically, reduce bias, and make responsible decisions that align with ethical standards in higher education [2]. Octaberlina, et al. [3] stated that AI can hinder the development of critical thinking and creativity by reducing active learning engagement and independent cognitive processes. Thus, critical thinking is more essential than ever for students to manage AI responsibly. The ability of large language models to transform teaching and learning in higher education is demonstrated by ChatGPT, a popular GenAI system [4]. It can be applied to critical-thinking tasks by offering real-time feedback on reasoning and argumentation [5]. Similarly, Nabilla, et al. [6] noted that Generative AI enhanced the flow of logical thinking, the structure of arguments, and the assessment of evidence, which are important for the development of critical thinking in learning English. However, as generative AI has grown, some have questioned whether it could compromise people's ability to develop critical thinking skills. Research indicates that both outcomes are feasible, as AI can assist with reasoning, but can also over-reliance if not utilized responsibly [3, 7]. Therefore, higher education must not only teach critical-thinking skills but also provide structured guidance on responsible AI use. Such training should involve targeted critical-thinking tasks that foster reflection and analytical engagement [8].

To address this challenge, Seufert and Rohwer [2], proposed a four-dimensional framework for critical thinking that includes analytical, evaluative, ethical reasoning and metacognitive skills. Although AI is widely used in education, limited evidence exists at the task level comparing how students perform with and without AI, and how they self-assess their use of AI to concerning ethical considerations. Thus, this study applied the four dimensions of the structured framework to explore when undergraduate students rely on AI, how this affects their results, how prompt quality influences learning outcomes, and how they rate themselves in terms of AI use. The research questions are as follows:

- Research Question 1: Which task involved the highest usage of AI by participants?
- Research Question 2: How are university students using AI while doing critical thinking tasks?

2) Theoretical background:

2.1 Critical thinking : Critical thinking includes essential skills such as analysis, evaluation, inference, explanation and self-regulation [9]. These skills enable learners to engage with information more critically, supporting deeper reasoning processes. Brookfield [10] describes critical thinking as the skills of categorizing assumptions, checking their validity, analyzing different perspectives, and producing informed decisions. Tarasova and Orel [11] mention that critical thinking involves analyzing arguments, evaluating evidence, categorizing assumptions, and reaching rational conclusions to support reasoned judgments. According to Indrašienė, et al. [12], critical thinking is widely regarded as a central goal of higher education. Assessment tools such as the Watson-Glaser Critical Thinking Appraisal (WGCTA) and the California Critical Thinking Skills Test (CCTST) operationalize these components through tasks like inference, deduction, interpretation, and evaluation [13, 14]. Beyond standardized tests, higher-order thinking tasks that ask students to propose hypotheses, evaluate arguments, and synthesize ideas have also been developed [15]. These approaches show the range of methods available to measure and foster critical thinking. However, scholars increasingly emphasize that fostering critical thinking in the age of GenAI requires moving beyond traditional assessments. Seufert and Rohwer [2] **framework is particularly relevant, as it integrates ethical reasoning and metacognition which are essential skills for responsible AI use.** Expanding this research with diverse populations and refined task designs will be vital for developing pedagogical frameworks that position AI as a tool for cultivating and assessing critical thinking rather than replacing it in higher education.

2.2 Generative Artificial Intelligence in Education. GenAI is seen mainly as enhancing, rather than replacing, educators, who provide uniquely human qualities such as creativity, judgment, and empathy [16, 17]. Benefits include personalized learning, adaptive feedback, and resource creation [18, 19]. It can also support metacognition by requiring students to evaluate outputs critically [20]. However, **studies mention over-reliance, cognitive dependency, and reduced independent problem-solving [21, 22]. AI "hallucinations" and fabricated references complicate academic integrity [23, 24], while bias, privacy, and access disparities raise further ethical issues [25, 26].** To mitigate these risks, scholars call for AI literacy and frameworks for responsible human-AI collaboration [27, 28]. Understanding this duality between AI as a supportive tool versus a cognitive crutch is central to our research, which seeks to provide empirical, task-level evidence on how students navigate these benefits and risks in practice.

2.3 Prompt Engineering in higher education. Prompt engineering is the practice of writing, improving and optimizing user inputs to generate desired outputs from generative AI systems [29, 30]. Prompt engineering is rapidly becoming a foundational skill in higher education, particularly as GenAI tools are integrated into teaching, research, and assessment practices. According to Aaron, et al. [31], prompt engineering is a developing capability that should be incorporated into student learning. A particular set of skills, and strategies is essential to achieve the most precise, relevant, and suitable responses [32]. Besides, effective use of generative AI depends on crafting prompts given to the AI systems [33]. To efficiently teach prompt engineering in higher education, structured pedagogical frameworks are vital. Lo [34] mentions the CLEAR framework to guide educators in designing effective prompts that are concise, logical, explicit, adaptive, and restrictive. Park and Choo [32] describe the PARTS framework to help educators design effective AI prompts, including five parts: personal, aim, recipients, theme, and structure. Skillful prompt engineering combined with critical assessment is vital for equipping students to interact effectively with AI technologies, as it supports higher-order thinking and metacognitive development, aligning with the goals of higher education [33]. Prompt engineering exercises should be designed to improve students' understanding of the subject matter but also develop critical thinking skills as they analyze and interpret the AI's output [35]. Thus, it is important to investigate the usage of prompts in critical thinking tasks and how students interact with generative AI ChatGPT based on the prompts they provide.

By analyzing the specific prompting strategies students employ, current research aims to uncover how different levels of engagement with AI, ranging from passive requests to active refinement, correlate with their performance on critical thinking tasks.

3) Methodology:

3.1 Participants. This study was conducted within an English-medium undergraduate program at a university in central Thailand. The participants were thirty students (N=30) enrolled in an English writing course focused on source critique. This population was particularly suitable for the study, as all participants had a demonstrated proficiency in English, with a minimum IELTS score of 6.0 across all four skills. The group comprised 13 males and 17 females, aged between 18 and 23, all of whom were first-year bachelor’s students. While the majority were Thai nationals (n=26), the sample also included students from China, Bangladesh, Indonesia, and the Philippines, adding cultural diversity. Participation was integrated into routine coursework, and students were made aware that their performance in the study would not affect their course grades.

3.2 Materials and Tasks. The study utilized seven critical thinking tasks adapted from a framework developed by Seufert and Rohwer [2] to assess critical thinking in the context of generative AI. These tasks were designed to measure skills across four key dimensions: analysis, evaluation, ethical reasoning (inference and decision-making), and metacognition. The sole AI tool permitted was ChatGPT-4o, which students accessed via the web interface. Student responses for each task were scored using the corresponding evaluation rubrics.

3.3 Procedure. The study protocol received Institutional Review Board (IRB) approval, and all participating students provided electronic informed consent prior to the session. The tasks were administered in a single, 120-minute session conducted during a scheduled class. Before beginning, participants received a comprehensive briefing on the study’s objectives and procedures. Each student worked individually, utilizing one continuous conversation thread in ChatGPT-4o to complete all seven tasks.

3.4 Data Analysis. A multi-faceted approach was used to analyse student data, focusing on performance, AI usage, and interaction patterns. First, student performance was quantified by summing the total score achieved on each task’s specific evaluation rubric. To measure AI reliance, two data sources were utilized: students’ self-reported contribution scores (0–100%) and the ChatGPT conversation logs. The log-based reliance was calculated by analyzing the percentage of text similarity between the AI’s responses and the student’s final submission. Building on this, academic transparency was operationalized as the absolute difference between a student’s self-reported estimate and the log-derived percentage. Finally, to understand user behaviour, prompting strategies were qualitatively coded from the chat logs into mutually exclusive categories describing the nature of the interaction, for example, *direct copy-paste*, *light edits/patchwriting*, *interlingual prompting*, etc. To ensure the reliability of this qualitative coding, the process was triangulated: three researchers independently coded the data, then met to resolve any discrepancies and reach consensus.

4) Results:

Q1: Which task involved the highest usage of AI by participants?

The initial research question sought to determine which critical thinking tasks prompted the greatest reliance on AI. As illustrated in **Figure 1**, a clear pattern emerged from the dialogue log analysis, indicating that AI reliance was strongly contingent on the task’s cognitive demands. To statistically validate this observation, a one-way ANOVA was conducted. The analysis confirmed a significant main effect of task type on AI reliance, $F(6, 203) = 14.687, p < .001$. The effect size was large ($\eta^2 = .303$), indicating that 30.3% of the variance in AI reliance was attributable to the specific task performed.

Figure 1. Mean Percentage of AI Contribution by Task



Post hoc analysis using Tukey’s HSD revealed significantly higher AI reliance for tasks involving inference and metacognition. Specifically, Task 4 (Inference - references formatting) showed the highest AI contribution (M = 79%), followed by the two Metacognition tasks, Task 7 (M = 72%) and Task 6 (M = 68%). These three tasks were statistically similar to one another and prompted significantly greater reliance on AI than the other tasks. In contrast, tasks focused on analysis and evaluation prompted substantially lower reliance on AI. Task 2 (Evaluation - Images) registered the lowest use (M = 11%), which was significantly lower than all other tasks except Task 1 (Analysis - Text). This confirms that students approached more direct, interpretative tasks with greater independence.

Q2: How are university students using AI while doing critical thinking tasks?

The second research question explored the relationship between how students used AI and their subsequent performance on critical thinking tasks. The analysis revealed that the specific strategy a student employed was a significant predictor of their success. From the chat logs and submissions, six distinct interaction strategies were identified, spanning from passive reception of AI output to active collaboration with the tool. The effectiveness of these strategies varied significantly, as shown in Table 1.

Table 1. AI Strategy Use and Mean Scores Relationship

Strategy	Mean Score	Description
Interlingual Mediation	67.60%	Using a native language to prompt and then translating/adapting the English output.
Iterative Refinement	66.00%	Modifying prompts and adjusting the AI’s output to fit the user’s needs.
Direct Reproduction	64.30%	Directly copying the prompt and pasting the AI’s unmodified answer.
Format-Based Refinement	54.20%	Asking the AI to make structural changes, such as shortening its answer.
AI-Led Clarification	46.50%	Prompting the AI to clarify its own output before using it.
No AI Used (Baseline)	40.30%	Completing the task without any assistance from AI.

Table 2 presents data that reveal distinct patterns in how students approach critical thinking tasks, linking their strategy choices directly to performance outcomes. First, on tasks requiring direct analysis and evaluation (Tasks 1, 2, and 5), students demonstrated a clear preference for working independently. In these cases, not using gen AI was the most common approach, peaking at 73.3% for Task 2. However, this independence did not lead to high performance; these tasks showed some of the lowest average scores in the study, such as 37.7% on Task 2. This suggests that while students avoid AI for interpretative work, they may struggle to complete these tasks successfully on their own.

Table 2. AI Strategy Use and Performance by Task

Strategy	Count	Percentage of Use
Task 1	Overall Score: 43.8%	
No AI Used	18	60%
Direct Reproduction	7	23.3%
Format-Based Refinement	2	6.7%
Iterative Refinement	1	3.3%
AI-Led Clarification	1	3.3%
Interlingual Mediation	1	3.3%
Task 2	Overall Score: 37.7%	
No AI Used	22	73.3%
Direct Reproduction	5	16.7%
Iterative Refinement	1	3.3%
AI-Led Clarification	1	3.3%
Interlingual Mediation	1	3.3%
Task 3	Overall Score: 53.0%	
No AI Used	18	60%

Direct Reproduction	6	20%
Iterative Refinement	4	13.3%
Interlingual Mediation	2	6.7%
Task 4	Overall Score: 26.3%	
Direct Reproduction	13	43.3%
No AI Used	8	26.7%
Format-Based Refinement	4	13.3%
Iterative Refinement	2	6.7%
AI-Led Clarification	2	6.7%
Interlingual Mediation	1	3.3%
Task 5	Overall Score: 53.9%	
No AI Used	17	56.7%
Direct Reproduction	7	23.3%
Format-Based Refinement	4	13.3%
Interlingual Mediation	2	6.7%
Task 6	Overall Score: 71.7%	
Direct Reproduction	14	46.8%
No AI Used	9	31%
Format-Based Refinement	4	13.8%
Interlingual Mediation	2	6.9%
Iterative Refinement	1	3.4%
Task 7	Overall Score: 80.0%	
Direct Reproduction	21	70%
No AI Used	5	16.1%
Format-Based Refinement	2	6.5%
Interlingual Mediation	2	6.5%

The study's "performance paradox" was most evident in the Task 4 referencing challenge. For this task, students had to use a non-standard citation style, which ChatGPT could not accurately reproduce because it was trained on standard formats. In response, participants defaulted to a passive the direct reproduction strategy and failed to verify the AI's flawed output critically. This uncritical reliance on the model led to the study's lowest success rate (26.3%), highlighting the significant risks of passive cognitive offloading.

In contrast to the above results, passive AI use and high performance was identified in the metacognitive assignments (Tasks 6 and 7). These tasks produced the study's highest success scores (71.7% and 80.0%, respectively), a result achieved despite a predominant student reliance on the direct reproduction strategy (used 46.8% and 70.0% of the time). This outcome suggests that for tasks emphasizing reflection and structure, the AI's capacity to articulate student ideas coherently aligns effectively with scoring rubrics, rendering even passive interaction a highly successful. The minimal use of active strategies, such as iterative refinement or interlingual mediation, reveals a critical skills gap. Students consistently avoided the more cognitively demanding "middle ground" of human-AI collaboration, opting instead for a simple binary choice: either do the work entirely alone or offload it passively to the AI.

The study also examined the accuracy of participants' self-reported AI contribution as an "integrity signal." The comparison between self-reported estimates and log data revealed a notable perception gap. While there were 204 instances of accurate calibration, there were also significant clusters of misestimation: 98 instances of exaggeration (over-reporting AI use) and 27 instances of underestimation (under-reporting AI use). This indicates that, while overall calibration was good, a substantial portion of interactions involved a mismatch between the perceived and actual AI contributions.

5) Discussion: Students use AI selectively, with high reliance on rule-based tasks like citation formatting and low reliance on nuanced interpretation, such as image analysis. Qualitative analysis shows a clear divide: successful students acted as "cognitive partners," using AI for refinement while maintaining critical oversight, whereas unsuccessful students treated it as a "cognitive replacement," offloading their thinking entirely. This supports the previous research linking cognitive offloading to weaker critical thinking, especially when AI is used to bypass "desirable difficulty" [36]. Ultimately, *how* students engage with AI is more predictive of success than *if* they use it. A "performance paradox" emerged in the citation task, where the highest AI reliance correlated with the lowest success rate (26.3%). This failure is twofold. First, LLMs are probabilistic systems, making them inherently unreliable for applying the strict, deterministic rules of a unique citation style. Second, students failed to evaluate the output critically, blindly trusting the AI. This aligns with broader research showing that LLMs often generate unsupported or fabricated citations [37, 38]. Success required active verification, underscoring critical evaluation as a core competency for students in the current AI-assisted work environment.

6) Implications for Teaching & Policy: According to this study, incorporating generative AI into higher education should affect the way critical thinking is taught, assessed, and guided by policy. Instructors need to help students build metacognitive skills and collaborate effectively with both peers and technology [20]. The findings caution against overreliance, especially in rule-based tasks like citation, where students performed the worst when accepting AI outputs unquestioningly. Thus, teaching practices and institutional policies should include AI use within a critical thinking framework. A key implication is the explicit integration of AI literacy into instruction. As Saddhono, et al. [39] note, AI can support critical thinking, but students must also learn to evaluate its outputs critically. Learning activities should include analyzing AI data, fact-checking evidence, and comparing outputs across tools or prompts to identify bias, inaccuracies, and hallucinations [40]. Additionally, the study found that mostly copy-pasting is associated with weaker outcomes and that incorporating process assessment practices could discourage passive copy-pasting. According to Cail [41], incorporating classroom assignments that require fact-checking generative AI outputs can enhance students' critical thinking skills, helping them develop a sharp eye for the quality, accuracy, and suitability of AI-generated content. Their study found that less experienced students in fact-checking, trusted generative AI outputs more and had more errors than those with more experience, who spot more inaccuracies and are cautious about trusting AI in the future. Likewise, Lamberti, et al. [42] mentioned about the structured critical thinking firstly to analyze AI-generated responses for correctness and coherence, then to critique the outputs by identifying the flaws such as factual errors, logical gaps, or hallucinations, and finally to revise or improve the AI's answers with their own reasoning. As a consequence, rather than final-product assignments, process assessment practices are needed in the classroom, such as incorporating fact-checked, structured critical thinking into the exercises. The role of prompting as a critical thinking exercise also needs to be emphasized. Prompting requires structuring the request to guide AI models to give the desired outcomes [43]. This process not only influences the quality of AI outputs but also requires students to engage in higher-order thinking by clarifying objectives, evaluating AI-generated responses, and fostering metacognitive awareness rather than using AI as a cognitive substitute [44, 45]. The prompting for critical exercises should include models such as CRAFT, which help students learn to scaffold their requests with precision, define context, adopt roles, and specify output formats that align with their goals [45]. The study also revealed a notable perception gap between students' self-reported estimates in human-AI collaboration. Subsequently, universities or institutes should develop clear disclosure requirements for AI use, explicitly distinguishing between supportive applications (e.g., brainstorming, rephrasing) and substitutive applications (e.g., direct reproduction of reasoning or output) to preserve the authenticity of student work while leveraging the benefits of AI [46]. This strategy is essential for addressing the disruptive effects of generative AI tools on evaluation procedures and for upholding academic integrity. [46, 47]. Institutions can promote transparency, allow educators to assess the degree of AI support in student work, and direct students toward the morally beneficial integration of AI as a learning tool rather than a substitute for their own intellectual work by requiring transparent reporting. [48]. Overall, these implications call for a critical thinking paradigm that balances the positions of Generative AI as a supportive tool for the correct usage while implanting safeguards that ensure authentic reasoning, ethical responsibility, and equitable participation in higher education.

Limitations and Future Work: While this study provides valuable insights into student-AI interaction, its findings are limited by context and sample size. Results may not generalize across disciplines, academic levels, or cultural settings. The cross-sectional design, based on a single 120-minute session, captures only a snapshot of behaviour, and the focus on ChatGPT-4o restricts applicability to other AI tools or future versions. To build upon this work, future research should aim to address these limitations. Studies could employ larger, more diverse samples across various institutions and academic disciplines to validate these initial

findings. Adopting longitudinal designs would be beneficial for tracking the development of students' human-AI collaboration and critical thinking skills over an entire semester or degree program. Additionally, intervention studies that assess the direct impact of pedagogical strategies, such as explicit instruction in prompt engineering using models like CRAFT [45], could provide causal evidence of their effect on student performance and ethical awareness.

7) **Conclusion:** This study provides a detailed empirical map of how undergraduate students utilize generative AI for critical thinking, revealing that the quality of engagement is a more powerful determinant of success than the quantity of AI use. The findings demonstrate that students rely most on AI for higher-order cognitive processes, such as inference, decision-making, and metacognition. A clear distinction emerged between effective and ineffective AI integration. Students who acted as "cognitive partners" by strategically using AI to filter, paraphrase, and refine their own thinking achieved stronger performance. Conversely, students who treated AI as a "cognitive replacement" through direct copy-pasting demonstrated weaker outcomes. This was starkly illustrated by a "performance paradox" in the citation task, where the highest AI reliance corresponded with the lowest success rate, a failure attributed to both the inherent unreliability of LLMs for rule-based tasks and students' lack of critical evaluation. Furthermore, the perception gap between students' self-reported and actual AI use highlights a significant ethical grey area, pointing to an urgent need for clear institutional policies on academic integrity and AI disclosure. Ultimately, this research calls for a pedagogical shift that positions AI not as a threat to be banned, but as a tool to be mastered. By embedding AI literacy, prompt engineering, and process-based assessment into the curriculum, higher education can foster an environment where students learn to collaborate with AI authentically and effectively, ensuring that technology supports rather than supplants human critical thinking.

8) **Acknowledgments:** This research was supported by the Swiss National Science Foundation (SNSF) for a joint project, "Next Generation of Digital Support for Fostering Participants' Academic Writing Skills: A Learning Support System based on Machine Learning (ML)." In this collaborative effort, the University of St. Gallen in Switzerland and Mahidol University in Thailand collaborated to investigate the influence of generative AI on undergraduate critical thinking.

9) References:

- [1] A. Johan, "An analysis of critical thinking skills through argumentative essay writing," *English Journal Literacy Utama*, vol. 8, no. 1, pp. 36-45, 2024.
- [2] S. Seufert and K. Rohwer, "Developing a framework for analysing and assessing critical thinking skills for the responsible use of generative AI in higher education," in *ICERI2024 Proceedings*, 2024: IATED, pp. 8579-8585.
- [3] L. R. Octaberlina, A. I. Muslimin, D. Chamidah, M. Surur, and A. Mustikawan, "Exploring the impact of AI threats on originality and critical thinking in academic writing," *Edelweiss Applied Science and Technology*, vol. 8, no. 6, pp. 8805-8814, 2024.
- [4] H. Gimpel et al., "Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide for students and lecturers," *Hohenheim Discussion Papers in Business, Economics and Social Sciences*, 2023.
- [5] S. Suh, J. Bang, and J. W. Han, "Developing critical thinking in second language learners: Exploring generative AI like ChatGPT as a tool for argumentative essay writing," *arXiv preprint arXiv:2503.17013*, 2025.
- [6] C. Nabilla, E. Apriani, and P. Gusmuliana, "Promoting students' writing critical thinking by using paragraph writing AI technology," Institut agama islam negeri curup, 2024.
- [7] M. Abbas, F. A. Jam, and T. I. Khan, "Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students," *International journal of educational technology in higher education*, vol. 21, no. 1, p. 10, 2024.
- [8] D. L. Jaramillo Gómez, A. J. Álvarez Maestre, A. E. Parada Trujillo, C. A. Pérez Fuentes, D. H. Bedoya Ortiz, and R. K. Sanabria Alarcón, "Determining Factors for the Development of Critical Thinking in Higher Education," *Journal of Intelligence*, vol. 13, no. 6, p. 59, 2025.
- [9] P. Facione, "A statement of expert consensus for purpose of educational assessment and instructions," *Research findings and Recommendations, American Philosophical Association (ERIC Document Reproduction Service No. ED 315 423)*, Newark, DE, 1990.
- [10] S. Brookfield, "Teaching for critical thinking," *International Journal of Adult Vocational Education and Technology (IJAVET)*, vol. 4, no. 1, pp. 1-15, 2013.
- [11] K. V. Tarasova and E. A. Orel, "Measuring students' critical thinking in online environment: Methodology, conceptual framework and tasks typology," *Вопросы образования*, no. 3 (eng), pp. 187-212, 2022.
- [12] V. Indrašienė et al., "The value of critical thinking in higher education and the labour market: The voice of stakeholders," *Social Sciences*, vol. 10, no. 8, p. 286, 2021.
- [13] I. Assessment, "California critical thinking skills test (CCTST)," ed: Retrieved, 2013.
- [14] D. K. Knox, "The California critical thinking skills test," *Clemson Thinks*, vol. 2, 2013.
- [15] S. Cargas, S. Williams, and M. Rosenberg, "An approach to teaching critical thinking across disciplines using performance tasks with a common rubric," *Thinking Skills and Creativity*, vol. 26, pp. 24-37, 2017.
- [16] C. K. Y. Chan and L. H. Tsi, "The AI revolution in education: Will AI replace or assist teachers in higher education?," *arXiv preprint arXiv:2305.01185*, 2023.
- [17] M. Louis and M. ElAzab, "Will AI replace teacher?," *International Journal of Internet Education*, vol. 22, no. 2, pp. 9-21, 2023.
- [18] M. Henderson et al., "Comparing Generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness," *Assessment & Evaluation in Higher Education*, pp. 1-16, 2025.
- [19] I. Pesovski, R. Santos, R. Henriques, and V. Trajkovic, "Generative AI for customizable learning experiences," *Sustainability*, vol. 16, no. 7, p. 3034, 2024.
- [20] P. Atchley, H. Pannell, K. Wofford, M. Hopkins, and R. A. Atchley, "Human and AI collaboration in the higher education environment: Opportunities and concerns," *Cognitive research: principles and implications*, vol. 9, no. 1, p. 20, 2024.
- [21] S. F. Ahmad et al., "Impact of artificial intelligence on human loss in decision making, laziness and safety in education," *Humanities and Social Sciences Communications*, vol. 10, no. 1, pp. 1-14, 2023.
- [22] J. Hutson, "Human-ai collaboration in writing: A multidimensional framework for creative and intellectual authorship," *International Journal of Changes in Education*, 2025.
- [23] J. Gravel, M. D'Amours-Gravel, and E. Osmanliu, "Learning to fake it: Limited responses and fabricated references provided by ChatGPT for medical questions," *Mayo Clinic Proceedings: Digital Health*, vol. 1, no. 3, pp. 226-234, 2023.
- [24] K. Masters, "Medical teacher's first ChatGPT's referencing hallucinations: Lessons for editors, reviewers, and teachers," *Medical Teacher*, vol. 45, no. 7, pp. 673-675, 2023.
- [25] M. Perkins, "Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond," *Journal of University Teaching and Learning Practice*, vol. 20, no. 2, pp. 1-24, 2023.
- [26] J. Qian, "Unleashing generative AI: funding implications and insights from China," *Journal of Asian Public Policy*, vol. 18, no. 2, pp. 431-454, 2025.
- [27] D. Long and B. Magerko, "What is AI literacy? Competencies and design considerations," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1-16.
- [28] A. Nguyen, Y. Hong, B. Dang, and X. Huang, "Human-AI collaboration patterns in AI-assisted academic writing," *Studies in Higher Education*, vol. 49, no. 5, pp. 847-864, 2024.
- [29] S. Ekin, "Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices," *Authorea Preprints*, 2023.
- [30] J. D. Velázquez-Henao, C. J. Franco-Cardona, and L. Cadavid-Higueta, "Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering," *Dyna*, vol. 90, no. SPE230, pp. 9-17, 2023.
- [31] L. Aaron, S. Abbate, N. Marae Allain, B. Fallon, and D. Gavin, *Optimizing AI in Higher Education: SUNY FACT? Guide*. State university of New York press, 2024.
- [32] J. Park and S. Choo, "Generative AI prompt engineering for educators: Practical strategies," *Journal of Special Education Technology*, vol. 40, no. 3, pp. 411-417, 2025.
- [33] D. Lee and E. Palmer, "Prompt engineering in higher education: a systematic review to help inform curricula," *International Journal of Educational Technology in Higher Education*, vol. 22, no. 1, p. 7, 2025.
- [34] L. S. Lo, "The CLEAR path: A framework for enhancing information literacy through prompt engineering," *The Journal of Academic Librarianship*, vol. 49, no. 4, p. 102720, 2023.
- [35] Y. Walter, "Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, p. 15, 2024.
- [36] M. Gerlich, "AI tools in society: Impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, p. 6, 2025.
- [37] M. Chelli et al., "Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: Comparative analysis," *Journal of medical Internet research*, vol. 26, no. 1, p. e53164, 2024.
- [38] W. Zhang, M. Aliannejadi, Y. Yuan, J. Pei, J.-H. Huang, and E. Kanoulas, "Towards fine-grained citation evaluation in generated text: A comparative analysis of faithfulness metrics," *arXiv preprint arXiv:2406.15264*, 2024.
- [39] K. Saddhono, R. Suhita, A. Rakhmawati, M. Rohmadi, and I. K. Sukmono, "AI and literacy: Developing critical thinking and analytical skills in the digital era," in *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 2024: IEEE, pp. 360-365.
- [40] L. Yingzhe, "Addressing "Hallucinations" in AI-generated content: Strategies for developing student fact-checking and information evaluation skills," *Artificial Intelligence Education Studies*, vol. 1, no. 2, pp. 48-62, 2025.
- [41] J. Cail, "Visualization of AI accuracy: A novel assignment for the teaching of critical thinking and science writing," *Teaching of Psychology*, vol. 52, no. 3, pp. 285-290, 2025.
- [42] W. F. Lamberti, S. R. Lawrence, D. White, S. Kim, and S. Abdullah, "Pilot study on generative AI and critical thinking in higher education classrooms," *arXiv preprint arXiv:2509.00167*, 2025.
- [43] P. Bansal, "Prompt engineering importance and applicability with generative AI," *Journal of Computer and Communications*, vol. 12, no. 10, pp. 14-23, 2024.
- [44] M. C. Murray, "Bridging the generative AI literacy gap: A guide to introducing prompt engineering in university courses," *Issues in Informing Science and Information Technology*, vol. 22, p. 010, 2025.
- [45] D. Siegle, "Using AI prompt engineering to improve gifted students' questioning," *Gifted Child Today*, vol. 48, no. 1, pp. 68-72, 2025.
- [46] B. L. Moorhouse, M. A. Yeo, and Y. Wan, "Generative AI tools and assessment: Guidelines of the world's top-ranking universities," *Computers and Education Open*, vol. 5, p. 100151, 2023.
- [47] C. Wu, H. Zhang, and J. M. Carroll, "AI governance in higher education: Case studies of guidance at big ten universities," *arXiv preprint arXiv:2409.02017*, 2024.
- [48] E. A. Ross and J. Baines, "Treading water: new data on the impact of AI ethics information sessions in classics and ancient language pedagogy," *Journal of Classics Teaching*, vol. 25, no. 50, pp. 181-190, 2024.