JUTLP
Journal of University Teaching & Learning Practice

The Impact of AI-Powered Text Generation on Undergraduate Students' Critical Thinking: A Study on ChatGPT Usage

Gohar Rahman[a],Eqab Ali A. Almutairi[b], Mohammad Almutairi[b], Badri Abdulhakim Mudhsh[c]
[a]Masaryk University, Czech Republic.
[b]The Public Authority for Applied Education and Training (PAAET), Kuwait.
[c]University of Technology and Applied Sciences, Salalah, Sultanate of Oman

Abstract

The adoption of generative AI (GenAI) tools in higher education has prompted questions about their impact on students' critical thinking. This study investigates the influence of ChatGPT on undergraduate cognitive development within a Pakistani university, using a mixed-methods quasi-experimental design (n=100 students, control v experimental). Over six weeks, the experimental group completed academic tasks aligned to Bloom's Revised Taxonomy using ChatGPT under guided conditions. Pre-and-post testing revealed significant improvements in lower-order skills - remembering, understanding and applying but no statistically significant improvement in higher-order skills - analysing, evaluating, and creating. These outcomes reflect the pedagogical design of the intervention, rather than inherent limitations of the tool. Thematic analysis indicated perceived benefits for higher-order thinking (idea generation, task structuring and evaluation) alongside concerns about over-reliance, hallucinations and integrity. We conclude that GenAI can scaffold lower-level learning when embedded into structured activities, whereas gains in higher-order thinking depend on deliberate instructional design. We recommend integrating AI-literacy training, explicit ethical guidelines, and assessment alignment to ensure critical thinking remains central in AI-enhanced curricula.

Editors

Section: Educational Technology
Senior Editor: A/Professor Rachel Fitzgerald
Associate Editor: Dr Jasper Roe

Publication

Submission: 25/4/25
Revised: 6/10/25
Accepted: 24/10/25
Published: 7/12/25

Keywords

Generative AI, critical thinking, cognitive offloading, higher education, AI ethics

Citation:

Rahman, G., Almutairi, E. A. A., Almutairi, M., & Mudhsh, B. A. (2025). The impact of AI-powered text generation tools on the critical thinking skills of undergraduate students. *Journal of University Teaching and Learning Practice, 21*(6). https://doi.org/10.53761/srkt0510

Introduction

The rapid acceleration of digital innovation and ubiquitous access to data have positioned technology as a transformative force in higher education (HE). This is especially relevant in cultivating essential cognitive competencies such as critical thinking. Critical thinking is defined as the ability to evaluate information, challenge assumptions, and generate logical, evidence-based solutions, it is a foundational skill for academic success and workplace readiness (Facione, 2020; Paul & Elder, 2020). As higher education evolves from content-heavy, discipline-specific instruction towards fostering transferable skills such as analytical reasoning, problem-solving, and creative synthesis (Pellegrino & Hilton, 2012; World Economic Forum, 2023), this shift aligns with demands of the global workforce (Dumitru & Halpern, 2023).

Artificial intelligence (AI), and specifically, generative AI (GenAI), is transforming traditional educational models. These technologies can produce text, media, and context-responsive outputs and are now increasingly embedded in higher education to enable personalised learning, support academic writing, simulate problem-solving processes, and reduce administrative workload (Al-Shahri & Mudhsh, 2025; Essel et al., 2024; Michel-Villarreal et al., 2023). Tools such as ChatGPT provide students with immediate feedback, example structures, and concise knowledge summaries that can scaffold content understanding and simulate creativity (Rudolph et al., 2023). However, this pedagogical shift also brings risks. Critics argue that the uncritical use of GenAI may encourage cognitive offloading, where students delegate higher-order thinking to machines, potentially undermining analytical depth, originality, and ethical reasoning (Cotton et al., 2023; O'Dea & O'Dea, 2023; Sweller, 2011). For instance, students who rely on AI-generated summaries may process content at surface level, diminishing engagement with synthesis, evaluation, and argument construction (Jin et al., 2023; Zhai & Wibowo, 2024;).

Recent empirical research attempts to reconcile GenAI's affordances with its potential risks by exploring synergistic frameworks, instructional design that trains students to use AI as a cognitive partner rather than a replacement (Rahimi, 2025). These models promote responsible and strategic use of GenAI to enhance critical thinking, communication, and domain-specific knowledge, skills, and attitudes (Tang et al., 2025). Tang et al.'s (2025) study, for instance, demonstrates that structured AI-literacy training can lead to significant gains in undergraduate students' ability to critique, revise, and integrate GenAI output into their own arguments. Similarly, Rahimi's (2025) framework advocates for AI-assisted reasoning through metacognitive scaffolding, ethical interrogation, and iterative writing tasks, highlighting how targeted training can mitigate risks while enhancing higher-order thinking. Nevertheless, such approaches remain underexplored in undergraduate contexts, particularly outside STEM or postgraduate environments (Essien et al., 2024; QAA, 2023).

To address this gap, the present study investigates the impact of ChatGPT, a widely used GenAI tool on undergraduate students' critical thinking development across multiple academic disciplines. Drawing on Bloom's Revised Taxonomy, which organizes cognitive development into progressive domains (remembering, understanding, applying, analyzing, evaluating, and creating), this study employs a quasi-experimental mixed-methods design (Krathwohl, 2002). A total of 100 undergraduate students were randomly assigned to control and experimental groups to compare AI-mediated and traditional approaches to academic task completion. The

experimental group engaged in guided interactions with ChatGPT, while the control group followed conventional instruction.

This research contributes to the literature in two ways. First, it expands empirical understanding of GenAI's influence on critical thinking by combining quantitative assessment with qualitative student reflections. Second, it advances recent calls for human-AI synergy, where technology is leveraged to enhance rather than replace cognitive rigor, ethical awareness, and learner agency (Hwang & Chang, 2023; Illinois College of Education, 2024; Rahimi, 2025; Tang et al., 2025). In doing so, it provides evidence-based recommendations for educators, institutions, and policymakers aiming to responsibly integrate GenAI in undergraduate curricula. This study addresses two pivotal questions:

1. How do AI text generators affect critical thinking skills among undergraduate students?
2. What challenges arise from using these technologies in developing such skills?

Literature Review

The Transformative Role of Generative AI in Education

Since ChatGPT's launch in November 2022, Generative AI (GenAI) tools have rapidly moved from novelty to essential components in modern education. These tools now support various functions, including grading automation, personalized learning, language feedback, summarization, and administrative efficiency (Al-Raimi et al. 2024; Davenport & Mittal, 2022; Moscardini et al., 2022; Ng, 2022). Their growing presence reflects a broader shift in higher education towards competency-based learning that prioritizes critical thinking, adaptability, and lifelong learning (World Economic Forum, 2023; Pellegrino & Hilton, 2012). Rahimi (2025) introduces the Critical Dialogic MKO model, a significant framework that reimagines GenAI as a Vygotskian More Knowledgeable Other (MKO). Unlike traditional passive learning tools, GenAI engages students in dialogic exchanges that promote justification, evidence-based reasoning, and exploration of multiple perspectives. Preliminary findings linked to this model suggest marked improvements in learners' critical thinking and epistemic vigilance. Building on this, Tang et al. (2025) offer the Dialogic Heteroglossia model, grounded in Bakhtin's theory of heteroglossia. This model encourages students to view AI-generated content as dialogic texts, prompting them to generate counterarguments, verify claims, and reframe ideas to gain deeper understanding. Empirical studies show that this interaction fosters metacognitive awareness, perspective-taking, and stronger argumentative skills. Together, these emerging frameworks underscore GenAI's transformative potential in shaping reflective, independent, and critically engaged learners through active dialogic learning.

Foundational Learning vs. Higher-Order Cognitive Gains

Recent meta-analyses indicate that Generative AI (GenAI) tools consistently support foundational academic skills, including paraphrasing, comprehension, and basic content assimilation (Albayati, 2024; Leiker et al., 2023; Lodge et al., 2023; Singh et al., 2025). However, their effectiveness in promoting higher-order cognitive skills, such as synthesis, evaluation, and critical analysis, remains inconsistent.

This discrepancy has been attributed to what scholars term "metacognitive laziness" (Zhai &

Wibowo, 2024), a phenomenon where students tend to accept AI-generated outputs without question. Such overreliance on automation can suppress learners' cognitive engagement, limiting opportunities for deep thinking. This concern aligns with earlier warnings about automation's potential to reduce the depth of student processing and critical reasoning (Gerlich, 2025; Sweller, 2011; Zawacki-Richter et al., 2019). To counteract these effects, educators are embedding metacognitive prompts within GenAI-supported tasks. Singh et al. (2025) demonstrated that questions like "How might this content mislead?" or "What alternative viewpoint exists?" prompted students to engage more critically, improving their analytical depth and argumentative coherence. Similarly, Wiboolyasarin et al. (2025) found that structured critical dialogues significantly enhanced learners' willingness to consider alternative perspectives and construct more complex arguments. These findings suggest that while GenAI excels at reinforcing basic academic skills, its potential to foster advanced cognitive outcomes depends heavily on how it is integrated into pedagogical design, particularly through tasks that actively encourage reflection, critique, and deeper inquiry.

Ethical, Epistemic, and Cultural Concerns

While generative artificial intelligence (GenAI) offers significant opportunities to enhance teaching and learning, its integration into higher education brings ethical, epistemic, and cultural challenges that demand critical attention. One key issue is AI "hallucination," where systems produce information that appears credible but is factually inaccurate. Research indicates students encounter such inaccuracies without recognising them (Lim et al., 2023; O'Dea & O'Dea, 2023), posing risks to academic integrity and diminishing students' epistemic trust in digital tools. To address this, Rahimi (2025) advocates a synergistic approach that encourages learners to engage critically with AI-generated texts by verifying facts, cross-checking claims, and interpreting responses within disciplinary contexts. Cultural bias represents another critical concern, as most GenAI models are trained on predominantly Western-centric datasets that perpetuate dominant ideologies and exclude alternative worldviews (Arora et al., 2023; Gallegos et al., 2023; UNESCO, 2024;). In response, UNESCO (2025) and Nemorin (2024) call for culturally inclusive AI design and pedagogies that promote epistemic diversity. Academic integrity further remains at risk, with studies revealing that over one third of students use chatbots for assessments without perceiving this as misconduct (Arora et al., 2023; Cotton et al., 2023; Peres et al., 2023). Collectively, these challenges underscore the need for robust AI literacy programs and clear institutional policies guiding responsible AI use.

Faculty Readiness and Instructional Integration

While students are rapidly embracing Generative AI (GenAI), many educators remain unsure about how to integrate these tools meaningfully into pedagogy. According to Ruediger et al. (2024), fewer instructors feel confident designing GenAI-based assignments that promote higher-order thinking. Much of the current training emphasizes technical skills, such as prompt engineering, rather than pedagogical strategies.

To bridge this gap, scholars recommend aligning GenAI use with established educational frameworks and dialogic teaching practices. Rahimi (2025) proposes a *synergistic pedagogy* that develops students' critical thinking and communication skills by combining a critical dialogic approach with competency-trained dialoguing using GenAI tools such as ChatGPT, positioned as

multidisciplinary more knowledgeable others. Tang et al. (2025) emphasize reimagining faculty roles, not as content transmitters, but as facilitators of inquiry-driven, AI-mediated learning experiences. In practice, faculty who embed reflective AI tasks in humanities courses report increased student ownership and more dynamic classroom discussions (Adhikari, 2025). Tan et al. (2024) further highlight the value of collaborative professional development, showing that faculty workshops focused on peer exchange significantly improve instructional integration of GenAI compared to isolated technical demonstrations.

Institutional culture also plays a vital role. McDonald et al. (2024) found that GenAI adoption is higher in universities that foster innovation through supportive policies, pedagogical training, and open dialogue. Overall, meaningful GenAI integration depends not only on individual faculty readiness but also on institutional commitment to pedagogical innovation and professional growth.

Emerging Empirical Evidence

Recent research from 2024–2025 provides compelling empirical evidence supporting the pedagogically informed integration of Generative AI (GenAI) in education. These studies emphasize the value of embedding AI use within dialogic, reflective, and collaborative learning environments to enhance student outcomes. Rahimi's (2025) implementation of the synergistic approach demonstrated notable improvements in students' critical thinking and communication skills, as reflected in their self-assessments and instructor observations. Similarly, Tang et al. (2025) examined the role of GenAI chatbots in science classrooms through the lens of heteroglossia, finding that dialogic engagement with AI strengthened students' reasoning, perspective-taking, and argumentation abilities—highlighting AI's potential to stimulate critical dialogue in STEM contexts.

Complementary studies further underscore the pedagogical benefits of structured AI use. Yu and Xie (2025) integrated AI-generated drafts with peer review and fact-checking exercises, reporting gains in students' argument construction and confidence in evaluating information credibility. Shen and Bai (2024) observed that learners engaged in AI-supported, dialogic writing curricula produced more coherent, reflective, and well-organized essays. Likewise, Francis et al. (2025) and Tan et al. (2024) found that AI-assisted collaboration enhanced students' thesis development, data critique, and synthesis of multiple viewpoints. Collectively, these findings suggest that when thoughtfully embedded within instructional design, GenAI can meaningfully advance students' critical reasoning and academic writing performance.

Positioning GenAI as a Cognitive Amplifier

To maximize the educational value of Generative AI (GenAI), educators must move beyond viewing it as a mere substitute for student effort and instead position it as a *cognitive amplifier*, a tool that enhances, rather than replaces, human reasoning (Rudolph et al., 2023). This shift ensures that learner agency remains central to the educational experience. One key strategy is threshold training, which involves establishing foundational domain knowledge before introducing GenAI tasks. This prevents students from relying on AI as a shortcut and reinforces essential disciplinary understanding (Rahimi, 2025). Dialogic prompting is another vital approach. Prompts that encourage critique, reflection, and questioning of AI outputs foster higher order thinking skills, such as analysis, synthesis, and evaluation (Tang et al., 2025; Singh et al., 2025).

In addition, metacognitive verification should be embedded into classroom practices. Teaching

students to assess the accuracy, bias, and credibility of GenAI-generated content cultivates epistemic vigilance and critical digital literacy (Rahimi, 2025; Leiker et al., 2023; Albayati, 2024). Finally, faculty development must go beyond technical training. Instructors need support in designing pedagogically aligned AI tasks that integrate seamlessly with disciplinary content, learning goals, and assessment practices (Cha et al., 2024). Together, these strategies position GenAI as a powerful ally in cultivating deeper learning and critical engagement

Research Gaps

Despite a surge in research exploring artificial intelligence (AI) in education, limited empirical attention has been paid to how AI text-generation tools influence the development of *critical thinking* among undergraduate students. While recent studies have documented the benefits of GenAI in supporting basic academic tasks such as summarization and comprehension (Lodge et al., 2023; Leiker et al., 2023), there remains a notable lack of investigation into its effects on advanced cognitive outcomes like synthesis, evaluation, and analytical reasoning (Tang et al., 2025; Singh et al., 2025). This oversight is especially significant given the increasing integration of GenAI in higher education and the centrality of critical thinking to academic success and professional readiness. Although emerging evidence suggests that structured, dialogic use of GenAI can foster argument quality, metacognitive awareness, and engagement with multiple perspectives (Rahimi, 2025; Francis et al. 2025; Tan et al., 2024), such findings are context-specific and require broader validation across disciplines and learning environments. As experts like Wilson and Holmes (2023) and Rahimi (2025) argue, it is imperative to investigate whether GenAI enhances or inhibits students' analytical autonomy and epistemic vigilance. Addressing this research gap will allow institutions to craft evidence-based policies and pedagogical practices that support critical thinking while ensuring responsible and meaningful use of AI in higher education.

Method

Research Design

This study employed a quasi-experimental mixed-methods design based on the convergent parallel model of Creswell and Plano Clark (2018), integrating quantitative and qualitative data to capture both measurable and experiential effects of AI text generators on undergraduate critical thinking skills. The quasi-experimental component involved pre-test and post-test measures for experimental and control groups, enabling comparison of cognitive skill gains. The mixed-methods approach triangulated these findings with qualitative data from reflective surveys and interviews, ensuring interpretations reflected both objective skill development and subjective learning experiences. This dual-focus design provided a comprehensive framework to evaluate how generative AI, specifically ChatGPT-3.5, mediates higher-order cognitive engagement among undergraduates.

The intervention lasted six weeks during Spring 2024 (March–May) at the National University of Modern Languages (NUML), Islamabad. Participants were divided into an experimental group, which engaged with AI-integrated assignments, and a control group, which completed equivalent tasks without AI support. Weekly assignments for the experimental group targeted Bloom’s Revised Taxonomy levels—remembering, understanding, applying, analyzing, evaluating, and

creating (Krathwohl, 2002). Tasks were scaffolded to promote active interrogation, critique, and refinement of AI outputs, while the control group completed identical tasks using traditional instruction.

Participants and Context

The sample comprised 100 undergraduate students across four disciplines—English Studies, Education, Computer Science, and Business Administration—ensuring balanced representation (25 students per discipline). Participants were aged 18–22, with no prior formal AI training, ensuring comparable baseline exposure and internal validity. Stratified randomization assigned 50 students to the experimental group and 50 to the control group, balancing gender, discipline, and prior academic performance (CGPA). Experimental group interactions with ChatGPT-3.5 were monitored via institutional logs, while the control group received standard pedagogy and a digital literacy workshop as a placebo control for motivational effects (Rudolph et al., 2023).

Intervention Protocol

The intervention consisted of two primary phases: a Week 1 AI Literacy Training module for the experimental group, followed by five weeks of structured task implementation across cognitive domains. The initial training was informed by current best practices in AI literacy, particularly focusing on the ethical and pedagogical implications of generative AI use in academic contexts (Cotton et al., 2023). The training emphasized four key competencies: prompt engineering, bias detection, hallucination mitigation, and academic integrity. It was delivered through a three-hour bilingual (English/Urdu) workshop, accompanied by handouts and a curated prompt repository tailored to Pakistani academic settings.

In the prompt engineering module, students learned how to craft discipline-specific and precise prompts, such as: “Generate a comparative analysis of cognitive theories using peer-reviewed psychological literature” (for Education) or “Explain recursive functions in Python with step-by-step examples” (for Computer Science). The bias detection segment trained students to identify and critique socio-cultural and gender biases in ChatGPT responses, using local examples such as rural versus urban development narratives in Pakistani economics. The hallucination mitigation component emphasized verifying AI-generated content against credible academic sources, including HEC research databases and peer-reviewed journals. Finally, the academic integrity module provided clear guidance on the responsible use of generative AI in scholarly work, including proper referencing (APA 7th), plagiarism avoidance, and critical reflection on AI-assisted writing (Cotton et al., 2023).

Instruments

To evaluate the intervention’s impact, the study employed two primary instruments: a Critical Thinking Skills Test and a Post-Intervention Mixed-Methods Survey, both of which were validated for content and construct reliability.

The Critical Thinking Test was a 34-item instrument developed in accordance with Bloom’s Revised Taxonomy (Krathwohl, 2002**)** and validated through expert review by three senior curriculum designers. It measured six dimensions: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. Sample items included: “List three principles of Keynesian theory” (Remembering), “Compare your explanation of recursive functions with ChatGPT’s”

(Understanding), “Debug the AI-generated Python code” (Applying), “Identify logical fallacies in ChatGPT’s climate change argument” (Analyzing), “Assess the ethical implications of ChatGPT’s suggestion on academic misconduct” (Evaluating), and “Develop a policy brief using AI ideas and original synthesis” (Creating). The test was pilot-tested with 15 students, yielding a Cronbach’s alpha of 0.83, suggesting high internal reliability. Variants of the test were adapted for each academic discipline to ensure contextual relevance.

The Mixed-Methods Survey, administered at the end of the intervention, included both quantitative Likert-scale items and qualitative open-ended questions (OEQs). The quantitative section consisted of 25 items measuring perceived learning gains, AI engagement patterns, and tool familiarity (e.g., “ChatGPT improved my ability to evaluate conflicting arguments: 1–5”). The qualitative section included five reflective prompts such as: “Describe how your understanding of topic X changed after using ChatGPT” and “Explain a moment where you challenged or disagreed with ChatGPT’s output.” Instrument development followed Nowell et al.’s (2017) criteria for qualitative trustworthiness, including peer-review, transparency in coding, and iterative revision.

Data Collection Procedure

Data collection occurred in four stages: Week 1 pre-test administration**,** Weeks 2–6 intervention activities**,** Week 6 post-test and survey administration, and post-study interviews with a subsample (n = 20). All ChatGPT interactions in the experimental group were archived for analysis of behavioural engagement, while the control group’s traditional outputs were retained for benchmarking. Data were anonymized, coded by participant ID, and stored on NUML’s secure servers following HEC Pakistan’s data protection protocols. Interviews were semi-structured and conducted in both English and Urdu, based on participant preference.

Data Analysis

Quantitative **data** from the critical thinking tests and survey responses were analysed using SPSS v28. Descriptive statistics (means, frequencies, standard deviations) were computed for overall performance, followed by paired-sample t-tests to measure within-group improvement and independent-samples t-tests to assess between-group differences. Effect sizes (Cohen’s d) were calculated to determine the magnitude of intervention effects, with thresholds set at 0.2 (small), 0.5 (medium), and 0.8 (large), in line with Field (2017). Additional analysis explored correlations between AI familiarity and post-test outcomes

Results

Quantitative

The study surveyed 100 undergraduate students aged between 18 and 22, the typical range for bachelor’s degree programs. The gender distribution was relatively balanced, with 52% identifying as male and 48% as female. When asked about their familiarity with AI text generation tools, 22% described themselves as "very familiar," 38% as "somewhat familiar," and 40% reported having no familiarity at all. Prior usage of such tools was also explored: 55% of students indicated they had previously used AI text generators, while 45% had not. These demographic insights provide a foundational understanding of the participants’ exposure to and engagement with AI

technologies prior to the study.

Pre-Test & Post-Test (Comparisons Across Bloom's Revised Taxonomy)

Table 1 provides a comparative overview of the pre-test and post-test scores for the experimental group across the six cognitive levels of Bloom's Revised Taxonomy. The data reveal statistically significant improvements in lower-order thinking skills, Remembering ($t = -7.12$, $p < 0.001$, $d = 1.42$), Understanding ($t = -6.85$, $p < 0.001$, $d = 1.32$), and Applying ($t = -4.20$, $p < 0.001$, $d = 0.85$), each demonstrating large effect sizes. These findings suggest that the instructional intervention was highly effective in enhancing foundational cognitive skills. In contrast, higher-order skills, such as Analyzing, Evaluating, and Creating, showed only minor gains, with p-values above 0.05 and negligible effect sizes, indicating no statistically significant change. This pattern implies that while the intervention successfully targeted Lower comprehension and application, it had limited impact on Higher cognitive tasks within the duration of the study.

Table 1

Pre-Test and Post-Test Scores Across Bloom's Revised Taxonomy (Experimental Group)

Cognitive Level	Pre-Test Mean (SD)	Cognitive Level	Pre-Test Mean (SD)	Cognitive Level	Pre-Test Mean (SD)
Remembering	3.10 (1.25)	4.60 (0.12)	−7.12	<0.001*	1.42
Understanding	3.05 (1.40)	4.45 (0.15)	−6.85	<0.001*	1.32
Applying	2.95 (1.30)	3.85 (0.18)	−4.20	<0.001*	0.85
Analyzing	2.80 (1.50)	3.10 (1.45)	−1.25	0.21	0.20
Evaluating	2.75 (1.60)	2.90 (1.55)	−0.85	0.40	0.10
Creating	2.60 (1.70)	2.75 (1.65)	−0.55	0.58	0.09

Note: $p < 0.05$ indicates statistical significance. Cohen's d thresholds: small (≥0.2), medium (≥0.5), large (≥0.8).

Statistical Inference

Analysis of Intervention Effects Using Paired-Sample t-Test

Table 2 presents the results of paired-sample t-tests examining the impact of the intervention across different cognitive domains of Bloom's Revised Taxonomy, based on item-level comparisons. The data reveal significant gains in Remembering (Q1–Q7), Understanding (Q2–Q8), and Applying (Q5–Q11), with all corresponding p-values below 0.001. These findings suggest the intervention was particularly effective in strengthening foundational cognitive skills. Conversely, no statistically significant improvements were observed in the higher-order domains of Analyzing (Q6–Q12), Evaluating (Q9–Q13), and Creating (Q10–Q14), as reflected by p-values of 0.21, 0.40, and 0.58 respectively.

Table 2

Paired-Sample t-Test Results Across Bloom's Revised Taxonomy

Cognitive Level	Test Pair	Pre-Test	Post-Test	t-value	p-value

		Mean	Mean		
Remembering	Q1-Q7	3.10	4.60	−7.12	<0.001*
Understanding	Q2-Q8	3.05	4.45	−6.85	<0.001*
Applying	Q5-Q11	2.95	3.85	−4.20	<0.001*
Analyzing	Q6-Q12	2.80	3.10	−1.25	0.21
Evaluating	Q9-Q13	2.75	2.90	−0.85	0.40
Creating	Q10-Q14	2.60	2.75	−0.55	0.58

Independent Sample t-Test for Group Comparisons

Table 3 compares post-test performance between the experimental and control groups using independent sample t-tests across Bloom's Revised Taxonomy. The data indicate statistically significant differences in favor of the experimental group in the lower-order cognitive domains, remembering ($t = 4.10$, $p < 0.001$), Understanding ($t = 3.75$, $p < 0.001$), and Applying ($t = 2.95$, $p = 0.004$). These results suggest that the intervention had a substantial impact on participants' acquisition of foundational knowledge and application skills. However, no significant differences were found in the higher-order domains of Analyzing, Evaluating, and Creating, with p-values well above the 0.05 threshold. This implies that while the experimental intervention effectively enhanced Lower cognitive skills, its influence on critical, evaluative, and creative thinking remained limited.

Table 3

Independent t-Test Results (Post-Test Scores)

Cognitive Level	Experimental Mean (SD)	Control Mean (SD)	t-value	p-value
Remembering	4.60 (0.12)	3.20 (1.30)	4.10	<0.001*
Understanding	4.45 (0.15)	3.15 (1.25)	3.75	<0.001*
Applying	3.85 (0.18)	3.00 (1.20)	2.95	0.004*
Analyzing	3.10 (1.45)	2.95 (1.50)	0.55	0.58
Evaluating	2.90 (1.55)	2.80 (1.60)	0.35	0.73
Creating	2.75 (1.65)	2.65 (1.70)	0.25	0.80

Qualitative results

Thematic analysis of the qualitative results was conducted to categorize feedback from the participants into two primary learning levels: Lower and Higher. These categories align with the cognitive domains of Bloom's Revised Taxonomy, where the Lower level pertains to remembering, understanding, and applying knowledge, while the higher level corresponds to analyzing and evaluating.

Table 04

Lower Level (Remembering, Understanding, Applying)

Sub-Theme	Description	Student Quotes

Proofreading	Students used ChatGPT to correct grammar and typos.	“I use ChatGPT to save time on routine tasks like organizing, editing, and formatting my writing." “It’s a helpful tool for checking and refining my written work.”
Structuring Assignments	Helped in creating outlines and organizing essays.	“When I come across unfamiliar topics, ChatGPT gives me a clear structure and overview, which I find very useful.” “I often rely on it to create assignment outlines.”
Summarizing Literature	Used to condense literature reviews to save time.	“Using AI helps offload repetitive tasks, giving me more time to think critically and creatively.” “AI can quickly handle large volumes of information, allowing me to concentrate more on interpreting and analyzing the material rather than just summarizing it.”

The Lower level is focused on foundational stages of learning, including remembering, understanding, and applying knowledge (see table 04). In this category, students primarily used ChatGPT as a tool to streamline routine academic tasks. A significant number of participants (33 out of 100) admitted to using ChatGPT to save time and assist with Lower academic tasks rather than enhance critical thinking skills. For example, 5 out of these 33 students utilized ChatGPT primarily for proofreading their work, ensuring that grammar and typographical errors were corrected. One student shared, " I use ChatGPT to save time on routine tasks like organizing, editing, and formatting my writing " Similarly, another participant stated, " It’s a helpful tool for checking and refining my written work " Furthermore, 16 students from this group found ChatGPT particularly helpful in structuring academic assignments, such as creating outlines for essays or dissertations. One student shared, " When I come across unfamiliar topics, ChatGPT gives me a clear structure and overview, which I find very useful"
Additionally, some participants (12 out of 33) highlighted that ChatGPT saved time when working on summarizing literature reviews. One participant explained, " Using AI helps offload repetitive tasks, giving me more time to think critically and creatively "AI can process a large amount of information quickly, helping me focus on analysis instead of spending time on summarizing literature."

Table 05
Higher Level (Analyzing, Evaluating)

Sub-Theme	Description	Student Quotes
New Perspectives	ChatGPT encouraged students to explore diverse viewpoints and challenge assumptions.	"ChatGPT opens the door to opinions I hadn’t considered before, it makes me pause and rethink my stance." "It nudges me to step outside my usual way of thinking and look at topics from angles I wouldn’t naturally explore."
Idea Generation	Used to brainstorm and develop	"It helps me kickstart my thoughts when I’m stuck, it’s like having a study partner to bounce ideas off."

	academic arguments.	"I use it to sketch out rough arguments, and then I refine them based on what I know and what I find through research."
Challenging AI Hallucinations	Students developed critical thinking by verifying AI-generated content.	"It's not always accurate, so I've learned to double-check facts and question what it generates." "I treat ChatGPT as a starting point, not a source, verifying its output has become second nature to me."

The higher level of cognitive engagement involves complex stages such as analyzing and evaluating, as defined in Bloom's Revised Taxonomy. In this study, 72 of 100 participants reported that ChatGPT played a significant role in enhancing their critical thinking. Students engaged with the tool in a self-directed manner, fostering reflection and deeper intellectual processing.

Three sub-themes emerged within this higher-level engagement: exposure to diverse viewpoints, academic brainstorming, and verification of information. Fifty-three students noted that ChatGPT introduced perspectives they had not previously considered, encouraging them to critically reassess their assumptions. One participant observed, "ChatGPT introduces unfamiliar perspectives that I wouldn't usually explore, and that challenges me to think more deeply about my arguments." Another noted, "It's like seeing the other side of the coin; it pushes me to weigh ideas more carefully."

Ten students identified academic brainstorming as a key benefit, particularly when encountering writer's block or difficulty structuring responses. As one student explained, "When I'm stuck, it gives me that initial push, and I build on those ideas using my own understanding." Another added, "It doesn't do the work for me, but it helps me organize my thoughts and map out my arguments more clearly."

The verification of information emerged as a third theme, with nine students reporting that critically evaluating AI-generated content strengthened their research-oriented thinking. One participant noted, "ChatGPT sometimes gets things wrong, so now I automatically double-check what it says." Another reflected, "I treat everything it produces as a draft; it forces me to question and validate each point before using it."

Although no students reported using ChatGPT to produce entirely original work, they consistently demonstrated advanced cognitive engagement by interpreting, refining, and integrating AI outputs into their academic reasoning, highlighting the tool's role in supporting higher-order thinking.

Table 06

Common Challenges Identified by Students

Challenge	Description	Student Quotes
Reliability and Accuracy	Concerns regarding the trustworthiness and timeliness of AI-generated information.	"Sometimes the information seems outdated or not entirely reliable, which makes me double-check everything." "I've noticed that it occasionally gives simplified answers that miss recent developments in the field."

Ethical Concerns	Worries about the reduction of human connection and the emotional disconnect in learning.	"Interacting with AI feels detached, there's no real conversation or emotional complex." "It's efficient, but it can't replicate the empathy or understanding you get in real discussions."
Digital Literacy	Difficulty in understanding how AI systems function and interpret input.	"It's confusing when the output doesn't reflect what I asked, it feels like a guessing game." "I still don't fully understand how it makes decisions or why it gives certain responses."
Bias and Oversimplification	AI-generated content may lack depth or reinforce narrow perspectives.	"It sometimes presents one-sided views, and that can limit how deeply I think about the topic." "If I rely too much on it, I feel like I'm not questioning things as critically as I should."

A thematic analysis of students' responses revealed four key challenges encountered when using AI text generators to support critical thinking development (see Table 06):

Reliability and Accuracy: Many students expressed concerns about the accuracy of the content produced by AI tools. Some noted that the information provided lacked complex or was not up to date. One student commented, "Sometimes, the answers feel too general or behind the curve, which makes me unsure about using them for academic work." Another noted, "I've had to fact-check multiple times because the tool gave me outdated or partially correct information." These responses point to a lack of confidence in the AI's ability to consistently deliver trustworthy content.

Ethical Concerns: Several participants raised ethical questions, particularly related to the loss of human interaction in learning environments. For instance, one student shared, "The experience feels robotic, it misses the human element that helps me emotionally connect with what I'm learning." Others emphasized that AI cannot replicate the social dynamics of peer discussion or mentorship. These concerns highlight the importance of maintaining human-centered learning alongside technological tools.

Digital Literacy: Several students reported difficulties navigating the functionality of AI tools. This challenge was especially pronounced when the tool's responses did not align with their expectations. One participant said, "I typed a clear question, but the response felt off, it didn't understand what I meant, and that was frustrating." Another added, "I still don't get how the AI processes what I write, so I don't always trust the output." These comments underscore a gap in digital literacy that can hinder effective use of such technologies.

Bias and Oversimplification: Concerns about biased or overly simplified outputs were also evident. Some students observed that AI tends to generalize complex issues or present a narrow set of perspectives. One remarked, "Sometimes it just scratches the surface, and I feel like I'm missing the deeper layers of the topic." Another reflected, "If I use it too much, I stop thinking critically because I'm just reacting to what it gives me." These insights suggest that while AI can assist with idea generation, overreliance may impede deeper analytical thinking.

In conclusion, while AI text generators offer new opportunities for learning, students face significant challenges related to accuracy, ethical concerns, digital literacy, and bias. Addressing

these issues is essential for effectively integrating AI tools into higher education and ensuring that they complement rather than compromise cognitive and critical development.

Discussion

This study explored the integration of Generative AI (GenAI) tools such as ChatGPT into undergraduate education through the lens of Bloom's Revised Taxonomy, revealing distinct patterns in cognitive engagement. The findings indicate that GenAI enhanced performance in lower-order domains, remembering, understanding, and applying, while yielding limited benefits for higher-order cognitive skills such as analyzing, evaluating, and creating. These results are broadly consistent with international trends, but also highlight important contextual and pedagogical differences that mediate the educational efficacy of AI tools.

The most immediate benefit of GenAI was its ability to scaffold lower-order cognitive tasks. Participants frequently reported that ChatGPT helped clarify abstract concepts, facilitated basic comprehension, and assisted in organising assignments. These outcomes align with prior research which has shown that students use AI tools to decode complex terminology and structure their writing more efficiently (Cano et al., 2023; Dwivedi et al., 2021; Shen & Bai, 2024; Yu & Xie, 2025). Cognitive load theory supports these findings, emphasising the value of reducing working memory demands to facilitate deeper learning (Essel et al., 2024; Hwang & Chang, 2023; Sweller, 2011; Tan et al., 2024).

Despite these foundational gains, the use of ChatGPT into more sophisticated cognitive tasks showed notable limitations. While some students attempted to use the tool for brainstorming, argument analysis, or creative exploration, the output often lacked disciplinary specificity and depth. This suggests that, without structured pedagogical scaffolding, AI tools may inadvertently reinforce surface-level engagement.

Contrastingly, studies that deliberately frame GenAI within critical pedagogical contexts report more promising results. For example, dialogic frameworks encourage students to interact with AI outputs as texts to be questions and verified, leading to enhanced perspective taking and argument quality (Rahimi 2025; Tang et al., 2025). The divergence between these studies and the present findings appears to stem from differences in instructional design, rather than limitations of GenAI itself.

The theme of "cognitive offloading" emerged strongly in this study, with many participants relying on ChatGPT to perform conceptual structuring, thereby reducing their own cognitive engagement. While some students developed critical skills by verifying AI-generated content, others incorporated inaccuracies without scrutiny. This inconsistency underscores the need for AI literacy training and explicit guidance on responsible use. Zhai and Wibowo (2024) have referred to this phenomenon as "metacognitive laziness," wherein learners accept AI outputs uncritically and bypass deeper reflection. This concern echoes earlier warnings by Sweller (2011) and has recently been reiterated by Gerlich (2025), who cautioned that unregulated AI use could erode intellectual rigor by promoting overdependence.

The study also uncovered instances where the limitations of AI served as a pedagogical advantage. In several cases, students identified hallucinations, fabricated or incorrect AI-generated information, which prompted them to cross-check sources and engage in deeper epistemic evaluation. This aligns with student-reported findings that learners often sharpen their

source scrutiny after encountering AI inaccuracies (Pitts et al., 2025). Similarly, Shen & Bai (2024) found that comparative analysis between AI-generated content and peer-reviewed sources enhanced students' analytical accuracy and source discernment.

However, these benefits were not uniformly distributed. Students with higher AI-literacy, those familiar with prompt engineering, source evaluation and reflective questioning, were better able to detect inaccuracies and engage critically. In contrast, those with limited AI awareness tended to accept outputs at face value. This pattern mirrors findings from intervention studies showing that overreliance on AI, especially among users lacking domain literacy, reduces critical engagement (Zhai et al., 2024), as well as broader educational investigations highlighting how students' self-efficacy and programming literacy influence reliance patterns (Pitts et al., 2025). These insights suggest that future implementations must prioritize training in epistemic awareness and critical prompt design.

A notable concern identified in this study was the inconsistency in students' responses to AI-generated hallucinations. While some participants engaged in verification and correction, others uncritically incorporated inaccurate information into their work. These divergent behaviours echo findings by Lim et al. (2023) and O'Dea and O'Dea (2023), who reported that a substantial proportion of students fail to recognize AI errors, often compromising academic outputs. Rahimi's (2025) "synergistic approach," which encourages learners to critically interrogate AI content, offers a promising instructional strategy to mitigate such risks and foster epistemic vigilance.

Cultural bias in GenAI outputs emerged as another critical theme. One student analyzing Nigerian literature observed that ChatGPT defaulted to Eurocentric interpretations, inadequately contextualizing postcolonial themes. This aligns with critiques by Gallegos et al. (2023) and Penkauskienė et al. (2019), who note that AI systems trained predominantly on Western-centric datasets can marginalize non-Western epistemologies. In response, UNESCO (2024) and Nemorin (2024) advocate for culturally responsive AI education, emphasizing inclusive training data and localized AI-literacy initiatives to address epistemic injustice. Without such interventions, GenAI risks perpetuating dominant narratives while sidelining diverse perspectives.

Academic integrity was also highlighted as an area of ambiguity. Although no direct plagiarism was reported, students expressed uncertainty about acceptable AI use, reflecting patterns documented by Cotton et al. (2023) and Peres et al. (2023). In the absence of clear institutional policies, learners may over-rely on GenAI for cognitive tasks, potentially crossing ethical boundaries unknowingly. Explicit academic guidelines combined with AI-literacy training can clarify expectations and promote responsible use.

A central implication is that the cognitive benefits of GenAI are not automatic but contingent on pedagogical design. Instructors who embedded AI within tasks emphasizing prompt formulation, dialogic interaction, and metacognitive monitoring observed higher levels of analytical and creative engagement, whereas superficial applications, such as formatting or rote reproduction, yielded minimal gains. These observations align with Tang et al. (2025) and Rahimi (2025), who stress the necessity of situating AI in reflective, critical tasks. Faculty preparedness is crucial; Ruediger et al. (2024) found only one-third of instructors felt confident designing AI-based activities targeting higher-order thinking, a gap reinforced by Adhikari (2025) and Tan et al. (2024).

The findings of this study call for a pedagogically grounded, blended approach to integrating tools like ChatGPT into undergraduate education. Students can utilize GenAI for low-stakes tasks such as generating research questions or creating initial essay outlines, while instructors should design

follow-up activities, like peer review, Socratic dialogue, and reflective writing, to foster critical thinking. As noted by Hwang and Chang (2023), such hybrid models enhance both efficiency and cognitive engagement. To ensure GenAI serves as a scaffold rather than a crutch, its use should be embedded in structured pedagogies such as Rahimi's (2025) dialogic MKO model and Tang et al.'s (2025) heteroglossia protocols. Incorporating metacognitive prompts, fact-checking exercises, and proper sequencing of foundational content before AI use is also essential. Moreover, AI literacy should be embedded into curricula, encompassing prompt engineering, ethical considerations, and bias awareness. Institutions must set clear usage policies, like those at the University of Illinois (2024), to promote transparency, academic integrity, and equitable access. Simultaneously, educators must receive ongoing professional development to integrate AI in ways that uphold student autonomy and intellectual growth (Southworth et al., 2023).

Conclusion

This study examined the cognitive impact of GenAI tools, particularly ChatGPT, on undergraduate students' critical thinking, using Bloom's Revised Taxonomy as an evaluative framework. The findings demonstrate that GenAI tools offer substantial support for foundational cognitive processes, such as remembering, understanding, and applying, but contribute less to higher-order skills like analysing, evaluating, and creating unless integrated through purposeful pedagogical design.

Students benefited most when using ChatGPT for conceptual clarification, assignment structuring, and information retrieval. These outcomes are consistent with cognitive load theory, which suggests that AI can reduce routine mental strain and free cognitive resources for deeper learning. However, when employed for complex tasks requiring interpretation or original thought, GenAI often produced shallow or formulaic outputs.

Crucially, the study found that the cognitive value of GenAI depends more on how it is embedded in learning environments than on the technology itself. Pedagogical frameworks that encourage dialogic interaction, critical verification, and epistemic reflection led to better engagement with higher-order thinking. Passive or uncritical use by contrast, can result in cognitive disengagement or dependency.

Challenges identified include inconsistent responses to AI-generated errors, cultural bias in content, and uncertainty around academic integrity. These issues highlight the need for AI-literacy education that fosters critical inquiry, cultural awareness, and ethical usage. GenAI tools hold promise as educational aids, but their effectiveness in developing critical thinking is contingent on deliberate, reflective integration. Educators and institutions must move beyond mere adoption and invest in strategies that complement human instruction, ensuring AI can serve as a cognitive partner in cultivating of independent, critical, and ethically aware learners.

Limitations and Future Research

While this study contributes valuable insights into the use of generative AI in undergraduate education, several limitations must be acknowledged. Firstly, the study was conducted at a single institution (NUML, Islamabad) with participants aged 18–22, which restricts the generalizability of the findings to broader educational or cultural contexts. Additionally, the intervention spanned a

short period and lacked sustained engagement with GenAI tools, limiting its capacity to evaluate long-term effects on learners' critical thinking skills and autonomy.

A key limitation is the absence of detailed descriptions of the pre- and post-test instruments and the specific tasks used during the intervention. This makes it difficult to determine whether the activities genuinely required higher-order cognitive engagement, as outlined by Bloom's Revised Taxonomy, or whether the assessments were appropriate for capturing such improvements. Although some qualitative themes indicated self-reported growth in analysis, evaluation, and reflection, these perceived gains were not reflected in the post-test scores. This discrepancy suggests a possible misalignment between the intervention's instructional design and the cognitive demands of the assessments.

Furthermore, the study did not engage in sufficient dialogue with contrasting findings in recent literature. For instance, Tang et al. (2025) reported measurable improvements in critical thinking when AI tools were embedded in scaffolded, inquiry-based tasks, an approach not fully replicated in this study. Similarly, Rahimi (2025) proposes a synergistic pedagogy for developing students' critical thinking and communication by combining a critical dialogic approach with competency-based dialoguing using GenAI tools such as ChatGPT, positioned as multidisciplinary more knowledgeable others. This may explain why mere exposure to ChatGPT-3 was insufficient for cognitive advancement in our study.

The use of ChatGPT-3, while accessible during the data collection phase, also represents a limitation. Its reasoning capabilities are more limited than newer models like GPT-4, which offer enhanced contextual understanding and responsiveness. Additionally, the study did not examine broader ethical or sociotechnical concerns, such as algorithmic bias or data privacy, nor did it address disparities in AI literacy and access, particularly relevant to contexts in the Global South.

Future research should consider adopting longitudinal, multi-institutional designs and employ richer assessment frameworks, such as the AAC&U VALUE Rubrics, to better capture interdisciplinary and metacognitive skills. Comparative studies across diverse educational systems, especially in underrepresented regions, are essential to uncover how infrastructural, cultural, and epistemic factors mediate AI's educational value. Importantly, future investigations must also consider the ethical dimensions of AI use in learning environments, fostering both critical digital literacy and equitable access.

Acknowledgements

The authors declare that they have no actual or perceived conflicts of interest. This research received no external funding beyond standard academic resources provided by the authors' respective institutions. Ethical approval for the study was granted by the Institutional Review Board (IRB) of the National University of Modern Languages (NUML), Islamabad, and all procedures involving human participants were conducted in accordance with institutional and national research ethics guidelines. Generative AI (GenAI) tools, specifically ChatGPT-3.5, were used solely as part of the instructional intervention and data collection for this study. Data supporting the findings of this study, including anonymized survey responses, test scores, and interview transcripts, are available from the corresponding author upon reasonable request.

References

Adhikari, E. R. (2025). Teaching with AI: A practical guide to a new era of human learning. Journal of Technical and Vocational Education and Training, 19(1), 112–114. https://doi.org/10.3126/tvet.v19i1.75730

Albayati, H. (2024). Investigating undergraduate students' perceptions and awareness of using ChatGPT as a regular assistance tool: A user acceptance perspective study. Computers and Education: Artificial Intelligence, 6, 100203. https://doi.org/10.1016/j.caeai.2024.100203

Al-Raimi, M., Mudhsh, B. A., Muqaibal, M. H., & Al-Yafaei, Y. (2024). To What Extent Has Artificial Intelligence Impacted EFL Teaching and Learning? A Systematic Review. JURNAL ARBITRER, 11(3), 399-412. https://doi.org/10.25077/ar.11.3.399-412.2024

Al-Shahri, H., Mudhsh, B, A. (2025). Generative AI in Modern Education: Exploring Teachers' Readiness, Benefits, and Challenges. Language Related Research, 16(5) :197-218. https://doi.org/10.48311/LRR.16.5.7

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cano, Y., Venuti, F., & Martinez, R. (2023). ChatGPT and AI text generators: Should academia adapt or resist? Harvard Business Publishing – Education. https://hbsp.harvard.edu/inspiring-minds/chatgpt-and-ai-text-generators-should-academia-adapt-or-resist

Cha, Y., Dai, Y., Lin, Z., Liu, A., & Lim, C. P. (2024). Empowering university educators to support generative AI-enabled learning: Proposing a competency framework. Procedia CIRP, 128, 256–261. https://doi.org/10.1016/j.procir.2024.06.021

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). 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

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). Sage Publications.

Dumitru, D., & Halpern, D. F. (2023). Critical thinking: Creating job-proof skills for the future of work. Journal of Intelligence, 11(10), 194. https://doi.org/10.3390/jintelligence11100194

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., &Eirug, A. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). Computers & Education: Artificial Intelligence, 6, 100198. https://doi.org/10.1016/j.caeai.2023.100198

Essien, A., Bukoye, O. T., O'Dea, X., &Kremantzis, M. (2024). The influence of AI text generators

on critical thinking skills in UK business schools. Studies in Higher Education, 49(5), 865–882. https://doi.org/10.1080/03075079.2024.2316881

Facione, P. A. (2020). Critical thinking: What it is and why it counts. Insight Assessment. http://dept.clcillinois.edu/vpe/gened/pdf/CriticalThinking_Facione.pdf

Field, A. (2017). Discovering statistics using IBM SPSS statistics (5th ed.). SAGE Publications. http://sro.sussex.ac.uk/id/eprint/77862/

Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in higher education: Balancing innovation and integrity. British Journal of Biomedical Science, 81.https://doi.org/10.3389/bjbs.2024.14048

Gerlich, R. N. (2025). Longitudinal effects of AI dependency on critical thinking. Journal of Higher Education, 96(3), 321–340.

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2023). Bias and Fairness in Large Language Models: A survey. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2309.00770

Hwang, G.-J., & Chang, C.-Y. (2023). A review of opportunities and challenges of chatbots in education. Interactive Learning Environments, 31(7), 4099–4112. https://doi.org/10.1080/10494820.2021.1952615

Illinois College of Education. (2024). AI in schools: Pros and cons. University of Illinois College of Education.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory Into Practice, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Leiker, D., Gyllen, A. R., Eldesouky, I., &Cukurova, M. (2023). Generative AI for learning: Investigating the potential of synthetic learning videos. arXiv Preprint.https://arxiv.org/abs/2303.12345

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., &Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. The International Journal of Management Education, 21(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. Australasian Journal of Educational Technology, 39, 1–8. https://doi.org/10.14742/ajet.8695

McDonald, N., Johri, A., Ali, A., & Hingle, A. (2024). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. arXiv.https://doi.org/10.48550/arxiv.2402.01659

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., &Gerardou, F. S. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. Education Sciences, 13(9), 856. https://doi.org/10.3390/educsci13090856

Nemorin, S. (2024). Towards decolonising the ethics of AI in education. Globalisation Societies and Education, 1 13 .https://doi.org/10.1080/14767724.2024.2333821

Ng, D. T. K., Leung, J. K. L., & Chu, S. K. W. (2022). AI literacy education for undergraduates: A scoping review. Computers & Education, 189, 104582. https://doi.org/10.1016/j.compedu.2022.104582

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to

meet the trustworthiness criteria. International Journal of Qualitative Methods, 16(1), 1–13.

O'Dea, X. C., & O'Dea, M. (2023). Is artificial intelligence really the next big thing in learning and teaching in higher education? A conceptual paper. Journal of University Teaching & Learning Practice, 20(5), 1–19. https://doi.org/10.53761/1.20.5.06

Paul, R., & Elder, L. (2020). Critical thinking: Concepts and tools (8th ed.). Rowman & Littlefield. http://pcwrgeography.pbworks.com/w/file/fetch/84314113/Concepts_Tools.pdf

Pellegrino, J. W., & Hilton, M. L. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. National Academies Press. https://www.hewlett.org/uploads/documents/Education_for_Life_and_Work.pdf

Peres, R., Schreier, M., Schweidel, D., &Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. International Journal of Research in Marketing, 40(2), 269–275. https://doi.org/10.1016/j.ijresmar.2023.03.001

Pitts, G., Rani, N., Mildort, W., & Cook, E. (2025). Students' reliance on AI in Higher Education: Identifying contributing factors. arXiv.org. https://arxiv.org/abs/2506.13845

QAA. (2023). Maintaining quality and standards in the ChatGPT era: QAA advice on the opportunities and challenges posed by generative artificial intelligence. Quality Assurance Agency for Higher Education. https://www.qaa.ac.uk/docs/qaa/members/maintaining-quality-and-standards-in-the-chatgpt-era.pdf

Rahimi, M. (2025). Synergising dialogic teaching with competencies-trained GenAI dialoguing for critical thinking and communication competencies. Journal of University Teaching and Learning Practice.https://doi.org/10.53761/vaa8h214

Ruediger, D., Blankstein, M., & Love, S. (2024). Generative AI and postsecondary instructional practices: Findings from a national survey of instructors. https://doi.org/10.18665/sr.320892

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching, 6(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Shen, B., & Bai, B. (2024). Enhancing Chinese university students' writing performance and self-regulated learning (SRL) writing strategy use through a strategy-based intervention. System, 122, 103249. https://doi.org/10.1016/j.system.2024.103249

Singh, A., Guan, Z., &Rieh, S. Y. (2025). Enhancing critical thinking in generative AI search with metacognitive prompts (Version 2). arXiv.https://doi.org/10.48550/arXiv.2505.24014

Southworth, J., Migliaccio, K., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers & Education: Artificial Intelligence, 4, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Sweller, J. (2011). Cognitive load theory. In Psychology of Learning and Motivation (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tan, X., Cheng, G., & Ling, M. H. (2024). Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence, 100355. https://doi.org/10.1016/j.caeai.2024.100355

Tang, K. S., & Putra, G. B. S. (2025). Generative AI as a dialogic partner: Enhancing multiple

perspectives, reasoning, and argumentation in science education with customized chatbots. Journal of Science Education and Technology.https://doi.org/10.1007/s10956-025-10240-1

UNESCO. (2024). What you need to know about UNESCO's new AI competency frameworks for students and teachers. https://www.unesco.org/en/articles/what-you-need-know-about-unescos-new-ai-competency-frameworks-students-and-teachers

UNESCO. (2025). Guidelines for culturally inclusive and ethical AI in education. https://www.unesco.org/reports/ai-education-2025

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., &Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. Ampersand, 100224.https://doi.org/10.1016/j.amper.2025.100224

Wilson, R., & Holmes, S. (2023). AI and cognitive skill development: A meta-analysis. Educational Research Review, 18(2), 200–215.

World Economic Forum. (2023). The future of jobs report 2023.https://www.weforum.org/reports/the-future-of-jobs-report-2023

Yu, H., & Xie, Q. (2025). Generative AI vs. teachers: Feedback quality, feedback uptake, and revision. Language Teaching Research Quarterly, 47, 113–137. https://doi.org/10.32038/ltrq.2025.47.07

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? International Journal of Educational Technology in Higher Education, 16(1). https://doi.org/10.1186/s41239-019-0171-0

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(1). https://doi.org/10.1186/s40561-024-00316-7