

GENERATIVE AI-ASSISTED ACADEMIC WRITING AND CRITICAL THINKING AMONG PAKISTANI UNIVERSITY STUDENTS: A MIXED-METHODS INVESTIGATION

Dr. Mohammad Tahir^{*1}, Farwa Hassan²

^{*1}Professor of Practice Department of Management Sciences, IIUI, Islamabad /Advocate High Court and
Federal Shariat Court

²Lecturer, Department of English

^{*1}drmtahirmalik2023@gmail.com, ²farwa.hassan@abasyn.edu.pk

DOI: <https://doi.org/10.5281/zenodo.23210244>

Keywords

Generative intelligence; Academic writing; Critical thinking; Cognitive Load Theory; AI evaluation and verification; Higher education

Article History

Received: 30 June 2026

Accepted: 21 September 2026

Published: 07 October 2026

Copyright @Author

Corresponding Author: *

Dr. Mohammad Tahir

Abstract

The rapid integration of generative artificial intelligence (GenAI) into higher education has transformed academic writing practices while raising concerns about its implications for students' critical-thinking development. This study investigated the relationship between GenAI-assisted academic writing and critical thinking among university students in Pakistan, with particular attention to whether GenAI functions as a cognitive scaffold or a substitute for independent reasoning. A sequential explanatory mixed-methods design was employed, comprising a quantitative survey followed by qualitative interviews. The quantitative phase examined AI-assisted idea generation, drafting and organization, language and revision, argument development, feedback and evaluation, and AI evaluation and verification in relation to critical thinking. The qualitative phase explored students' experiences of productive AI use, critical evaluation, and cognitive dependence. The findings indicated significant positive relationships between GenAI-assisted writing dimensions and critical thinking, with AI evaluation and verification emerging as the strongest predictor ($\beta = .39, p < .001$), followed by AI-assisted feedback and evaluation ($\beta = .22, p < .001$) and argument development ($\beta = .19, p < .001$). The qualitative findings further indicated that GenAI supported critical thinking when students used it for feedback, alternative perspectives, argument refinement, and verification, whereas excessive delegation of cognitive tasks encouraged dependence and reduced independent reasoning. The findings extend Cognitive Load Theory by distinguishing between productive cognitive-load reduction and cognitive offloading. The study concludes that the educational value of GenAI depends less on its availability than on how critically and purposefully students interact with it. Universities

should therefore emphasize AI literacy, source verification, reflective AI use, process-based assessment, and responsible human–AI collaboration.

INTRODUCTION

The rapid development of generative artificial intelligence (GenAI) has fundamentally altered the way knowledge is produced, communicated, and evaluated in higher education. Unlike earlier educational technologies that primarily supported information retrieval, automated assessment, or adaptive learning, contemporary GenAI systems based on large language models (LLMs) can generate coherent academic prose, summarize complex material, suggest arguments, revise language, provide explanations, formulate questions, and interact with students through iterative dialogue. Since the emergence of ChatGPT and comparable systems, universities have increasingly confronted both the pedagogical opportunities and the intellectual risks associated with AI-assisted learning and academic writing (Kasneci et al., 2023; Qian, 2025). Recent systematic evidence indicates that GenAI is being incorporated into higher education for writing support, personalized feedback, idea generation, learning assistance, and higher-order cognitive activities, although the outcomes depend substantially on how the technology is pedagogically integrated (Daniel et al., 2025; Qian, 2025).

Academic writing is particularly important in this debate because writing is not merely a mechanism for communicating already-developed knowledge; it is also a cognitive process through which students organize ideas, interpret evidence, construct arguments, compare perspectives, identify inconsistencies, and develop reasoned conclusions. Consequently, when students use GenAI to generate or substantially restructure academic text, the technology may influence not only the quality and efficiency of their writing but also the cognitive processes through which academic reasoning develops. GenAI can potentially function as a cognitive scaffold by providing alternative explanations, questioning assumptions, generating counterarguments, identifying weaknesses in an argument, and offering immediate feedback. Such applications may encourage students to revise, evaluate, and refine their reasoning rather than simply reproduce information (Kasneci et al., 2023; Daniel et al., 2025). Recent evidence from experimental research also suggests that appropriately designed ChatGPT interventions can improve academic performance and higher-order thinking, although the magnitude and durability of these effects remain context dependent (Deng et al., 2025).

At the same time, the educational value of GenAI cannot be inferred simply from the quality of the text it produces. A well-written AI-generated essay may conceal limited student engagement with the underlying reasoning process. When students accept AI-generated arguments without checking evidence, questioning assumptions, verifying references, or comparing alternatives, GenAI may encourage cognitive offloading and reduce opportunities for independent intellectual engagement. Recent research therefore identifies a fundamental duality in the relationship between GenAI and critical thinking: the technology can stimulate analysis and reflection when students actively interrogate its outputs, but it can undermine critical engagement when it becomes a substitute for students' own reasoning (Liu et al., 2025; Qian, 2025). A recent systematic review of GenAI and critical thinking in higher education similarly found evidence of both cognitive benefits and risks, with positive effects being more apparent

when AI use was accompanied by structured prompts, reflection, evaluation, and pedagogical guidance (Liu et al., 2025).

Critical thinking is particularly relevant because university education is expected to develop students who can analyze information, evaluate evidence, identify assumptions, make reasoned inferences, and reach defensible conclusions. The increasing availability of GenAI changes the nature of these competencies rather than eliminating their importance. Students now need not only to produce academic arguments but also to evaluate AI-generated arguments, distinguish credible information from plausible-sounding inaccuracies, identify unsupported claims, recognize bias, and determine when an AI response requires correction or further verification. Consequently, critical AI literacy and conventional critical-thinking skills are becoming increasingly interconnected. Recent research on artificial intelligence and critical thinking in higher education identifies uncritical dependence on AI, digital-literacy disparities, limited transparency of AI systems, and inadequate institutional guidance as major challenges to maintaining students' evaluative and self-regulatory capacities (2025).

The academic-writing dimension is equally consequential. GenAI can assist students with brainstorming, outlining, grammar, vocabulary, coherence, paraphrasing, summarization, and revision. These functions may be especially valuable for students who experience difficulties with academic English or who require additional support in transforming ideas into academically acceptable prose. However, excessive dependence on AI-generated text may weaken students' ownership of arguments and reduce the amount of effort devoted to constructing and defending their own findings. This issue is especially important in multilingual higher-education environments where English is frequently used as the language of academic publication and instruction. GenAI therefore creates a paradox: the same technology that can lower linguistic barriers and facilitate academic writing may also reduce students' opportunities to practice the cognitive processes embedded within writing.

The Pakistani higher-education context provides an important setting for examining this paradox. Pakistani universities are experiencing increasing student exposure to ChatGPT and other AI-based writing technologies, while institutional policies, instructional practices, and students' AI-literacy levels continue to develop. Recent Pakistan-specific research has begun to examine ChatGPT's implications for academic writing, learning, academic integrity, and critical thinking. For example, recent empirical work involving Pakistani university students has reported both perceived benefits and concerns surrounding ChatGPT use, particularly with respect to academic learning and integrity. Other emerging studies have specifically examined ChatGPT-assisted academic writing and critical thinking among Pakistani undergraduate students, suggesting that guided use may produce more favorable cognitive outcomes than unguided use.

However, the Pakistani evidence base remains comparatively limited and fragmented. Existing studies tend to concentrate on students' perceptions, general acceptance of ChatGPT, academic performance, writing proficiency, or academic-integrity concerns, while fewer studies systematically examine how different forms of GenAI-assisted academic writing influence critical-thinking development. This distinction is important because simply measuring frequency of AI use cannot reveal whether students are using GenAI as a learning scaffold, an editing assistant,

an idea-generation tool, or a substitute for independent academic reasoning. The cognitive consequences of these different forms of use are unlikely to be identical.

Recent international research reinforces this concern. A systematic review of 158 studies on GenAI and academic-skill development reported substantial evidence of benefits for cognitive and technical skills but emphasized the need for responsible use, continued monitoring, and stakeholder engagement (Daniel et al., 2025). Similarly, a systematic review of pedagogical applications in higher education identified growing interest in creativity, critical thinking, learner autonomy, and prompt literacy, while simultaneously warning that overreliance on GenAI may outsource critical cognitive and metacognitive processes (Qian, 2025). Experimental evidence further indicates that ChatGPT can improve higher-order thinking propensities, but researchers have called for objective measures and longer-term research rather than relying exclusively on students' self-reported perceptions (Deng et al., 2025).

The methodological approach is therefore important. A mixed-methods investigation can provide a more comprehensive explanation than either quantitative or qualitative research alone. Quantitative analysis can determine whether GenAI-assisted academic writing is statistically associated with students' critical-thinking capabilities and can examine the strength and direction of such relationships. Qualitative evidence, in contrast, can illuminate how students actually use GenAI during academic writing, how they evaluate AI-generated content, why they accept or reject AI suggestions, and whether they perceive GenAI as a tool for learning or as a substitute for independent work. Combining these forms of evidence can therefore distinguish between AI-assisted writing that supports cognitive development and AI-dependent writing that potentially displaces critical engagement.

Accordingly, this study investigates the relationship between generative AI-assisted academic writing and critical thinking among Pakistani university students through a mixed-methods approach. Rather than treating GenAI use as a single undifferentiated activity, the study conceptualizes academic writing assistance as a multidimensional process involving idea generation, drafting, language refinement, argument development, feedback, and revision. It further examines whether students' patterns of engagement with AI—particularly verification, reflection, evaluation, and independent judgment—help explain variations in critical-thinking outcomes. The study consequently addresses an important contemporary question in higher education: whether GenAI should function primarily as a replacement for students' academic writing labor or as a cognitive partner that strengthens students' capacity to reason, evaluate, and construct knowledge independently.

Problem Statement

The emergence of generative artificial intelligence has created a fundamental challenge for higher education: universities must determine how to exploit the educational benefits of AI-assisted academic writing without weakening the intellectual capabilities that higher education is intended to develop. GenAI tools can generate academically structured text within seconds, provide instant feedback, improve language accuracy, suggest alternative arguments, and assist students in organizing ideas. These capabilities have substantial pedagogical value, particularly for students who face difficulties with academic writing. Nevertheless, the increasing ability of students to outsource substantial portions of the writing process to GenAI raises a critical

question regarding what happens to the cognitive activities embedded in academic writing when the technology performs them on students' behalf.

The central problem is not simply whether university students use GenAI, but how the manner and intensity of that use influence their critical-thinking development. Academic writing traditionally requires students to engage in multiple higher-order cognitive processes, including identifying and interpreting relevant information, comparing competing perspectives, evaluating evidence, constructing arguments, recognizing assumptions, synthesizing literature, and drawing logical conclusions. If GenAI is used to generate ideas, arguments, paragraphs, and conclusions without substantive student evaluation, some of these cognitive activities may be transferred from the learner to the technological system. Such cognitive offloading could produce academically polished outputs while simultaneously limiting opportunities for students to develop independent reasoning and evaluative judgment.

Conversely, portraying GenAI as inherently detrimental to critical thinking would also be theoretically and empirically inadequate. GenAI can provide students with alternative perspectives, counterarguments, explanations, formative feedback, and opportunities to test their own reasoning. When students are required to critique AI-generated responses, verify evidence, identify factual inaccuracies, improve weak arguments, and justify revisions, interaction with GenAI may become a form of cognitive scaffolding rather than cognitive substitution. Recent systematic reviews indicate precisely this dual effect: GenAI can support critical-thinking development when its use is structured around reflection and evaluation, while uncritical or unguided use may encourage dependence and reduce intellectual autonomy (Liu et al., 2025; Qian, 2025). Thus, the educational problem involves understanding the conditions under which GenAI-assisted writing strengthens rather than displaces critical thinking.

This issue is particularly significant in Pakistan, where higher education operates within a multilingual academic environment and where students increasingly encounter English-language academic writing requirements. For many Pakistani university students, GenAI may provide valuable assistance with grammar, vocabulary, coherence, paraphrasing, organization, and idea development. However, these same functions can create dependency if students increasingly rely on AI to formulate arguments and express ideas that they have not independently developed. Recent Pakistan-specific research has begun to document students' use of ChatGPT, concerns about academic integrity, and potential effects on writing and critical thinking, but the emerging literature remains insufficiently developed and fragmented.

A further problem concerns the measurement of GenAI use. Much existing research treats AI use as a broad or frequency-based variable, such as whether students use ChatGPT and how often they use it. Such measures provide limited insight into the educational consequences of AI because two students may use the same technology in fundamentally different ways. One student may use GenAI to obtain an initial explanation, challenge the response, verify sources, identify weaknesses, and independently rewrite the argument. Another may copy the generated response with minimal evaluation. Although both students can be classified as "GenAI users," their cognitive engagement is substantially different. Research must therefore move beyond simple measures of adoption and examine the nature, purpose, and quality of AI-assisted academic-writing practices.

Another important gap concerns the limited integration of quantitative and qualitative evidence. Quantitative studies can establish associations between GenAI use and critical-thinking outcomes but may not adequately explain the cognitive processes underlying those relationships. Conversely, qualitative studies can reveal students' experiences, motivations, and concerns but may provide limited evidence regarding the magnitude or statistical significance of relationships between variables. A mixed-methods approach is consequently appropriate because it can establish empirical relationships while also explaining how and why students use GenAI during academic writing.

The existing international literature also demonstrates methodological limitations. Although recent systematic reviews and meta-analyses suggest that GenAI and ChatGPT can positively influence academic performance and higher-order thinking under certain conditions, researchers have emphasized substantial contextual and methodological heterogeneity, including differences in intervention duration, disciplinary setting, assessment methods, instructional design, and student characteristics (Deng et al., 2025; Daniel et al., 2025). Furthermore, much of the emerging evidence comes from contexts outside Pakistan. Findings generated in technologically mature or institutionally different higher-education systems cannot automatically be generalized to Pakistani universities because differences in language, pedagogical practices, digital access, institutional AI policies, assessment systems, and students' levels of AI literacy may shape the effects of GenAI.

The research gap is therefore threefold. First, there is insufficient empirical evidence regarding the relationship between GenAI-assisted academic writing and critical thinking among Pakistani university students. Second, existing research does not adequately distinguish between productive, reflective use of GenAI and passive or substitutive use that may encourage cognitive offloading. Third, there is limited mixed-methods evidence capable of integrating statistical relationships with students' actual experiences, reasoning processes, verification practices, and perceptions of AI-generated academic content.

This gap has important implications for universities, instructors, curriculum developers, and policymakers. If GenAI is integrated without appropriate pedagogical safeguards, students may become increasingly dependent on automated text generation while receiving insufficient opportunities to practice independent analysis and argumentation. Conversely, overly restrictive institutional responses may prevent students from developing the AI literacy and critical evaluation skills required for contemporary academic and professional environments. UNESCO has consequently advocated a human-centred approach emphasizing ethical, safe, meaningful, and pedagogically appropriate use of GenAI in education (Miao & Holmes, 2023).

The present study addresses this problem by investigating generative AI-assisted academic writing and critical thinking among Pakistani university students using a mixed-methods design. It seeks to move beyond the simplistic question of whether GenAI is "good" or "bad" for students and instead examine how GenAI-assisted writing practices, patterns of engagement, evaluation, verification, and independent judgment are associated with critical-thinking development. By addressing this problem within the Pakistani higher-education context, the study aims to provide empirically grounded evidence for designing AI-supported academic-writing practices that preserve intellectual autonomy while exploiting the educational potential of generative technologies.

Research Questions

1. What is the relationship between generative AI-assisted academic writing and critical thinking among Pakistani university students?
2. To what extent does the use of GenAI for idea generation, drafting, language refinement, argument development, and revision influence students' critical-thinking skills?
3. How does students' critical evaluation and verification of AI-generated content affect the relationship between GenAI-assisted academic writing and critical thinking?
4. What are the perceived benefits and challenges of GenAI-assisted academic writing for students' analytical, evaluative, and reasoning abilities?
5. How do students describe their experiences of using GenAI as a cognitive support tool versus a substitute for independent academic writing?
6. How do the quantitative findings regarding GenAI-assisted academic writing and critical thinking correspond with students' qualitative experiences and reported patterns of AI use?
7. What pedagogical and institutional strategies can Pakistani universities adopt to promote responsible GenAI-assisted academic writing while strengthening students' critical-thinking capabilities?

Research Objectives

1. To examine the relationship between generative AI-assisted academic writing and critical thinking among Pakistani university students.
2. To determine the effects of GenAI-assisted idea generation, drafting, language refinement, argument development, and revision on students' critical-thinking skills.
3. To assess the role of students' critical evaluation and verification of AI-generated content in shaping the relationship between GenAI-assisted academic writing and critical thinking.
4. To explore students' perceived benefits and challenges associated with GenAI-assisted academic writing, particularly regarding analysis, evaluation, reasoning, and independent judgment.
5. To investigate how Pakistani university students distinguish between using GenAI as a cognitive-support tool and using it as a substitute for independent academic writing.
6. To integrate quantitative and qualitative evidence to explain how and why GenAI-assisted academic-writing practices influence students' critical-thinking development.
7. To develop evidence-based pedagogical and institutional recommendations for integrating GenAI into academic writing while preserving and strengthening students' critical-thinking capabilities.

Significance of the Study

Theoretical Significance

The study contributes to the emerging theoretical literature on artificial intelligence in higher education by conceptualizing GenAI-assisted academic writing as a potentially dual-purpose cognitive technology. Rather than assuming that AI use automatically improves or diminishes learning, the study distinguishes between AI use that supports cognitive scaffolding and use that encourages cognitive substitution or offloading. This perspective can strengthen contemporary explanations of how digital technologies interact with higher-order thinking and academic

learning. The study also contributes to the growing literature connecting generative AI, AI literacy, metacognition, academic writing, and critical thinking.

Empirical Significance

The study provides empirical evidence from Pakistani universities, a context that remains comparatively underrepresented in the international GenAI literature. By examining both measurable relationships and students' lived experiences, the study can provide a more comprehensive understanding of the cognitive consequences of AI-assisted academic writing. Its mixed-methods design is particularly valuable because quantitative evidence can establish the strength and direction of relationships, while qualitative evidence can explain students' motivations, verification practices, perceptions, and patterns of dependence or critical engagement.

Practical Significance

For university instructors, the findings can help identify productive ways of incorporating GenAI into academic-writing instruction. Instead of using AI merely for text production, instructors may design activities requiring students to critique AI-generated arguments, verify sources, identify inaccuracies, compare alternative responses, improve weak reasoning, and justify their revisions. Such practices can transform GenAI from a writing substitute into a tool for reflective learning and critical engagement.

The study can also assist students in understanding responsible AI-assisted academic writing. The findings may encourage students to retain intellectual ownership of their work by independently developing arguments, checking AI-generated claims, verifying references, recognizing potential biases, and treating AI output as material for evaluation rather than as an unquestionable source of knowledge.

Institutional Significance

For Pakistani universities, the study can provide evidence for developing institutional guidelines concerning acceptable and responsible GenAI use in academic writing. Rather than adopting exclusively permissive or prohibitive approaches, universities can develop differentiated policies that recognize legitimate applications such as brainstorming, language improvement, formative feedback, and revision while establishing clear expectations regarding authorship, transparency, verification, and independent intellectual contribution.

Policy Significance

The study can inform higher-education policymakers and regulatory bodies regarding the development of AI-literacy frameworks, academic-integrity policies, assessment guidelines, and faculty-development initiatives. The evidence may support policies that emphasize human-centred and pedagogically meaningful AI integration, consistent with international recommendations for responsible GenAI use in education (Miao & Holmes, 2023).

Significance for Future Research

Finally, the study can establish a foundation for future longitudinal, experimental, and comparative research. Subsequent studies may examine whether prolonged GenAI use produces sustainable changes in critical thinking, whether different patterns of prompting produce different cognitive outcomes, and whether effects vary according to discipline, academic level, language proficiency, AI literacy, or socioeconomic background. The study can therefore contribute to a developing research agenda concerned not merely with whether students use GenAI, but with how, why, and under what conditions GenAI contributes to meaningful intellectual development.

Literature Review

1. Generative Artificial Intelligence in Higher Education

Generative artificial intelligence (GenAI) has rapidly transformed higher education by enabling students to generate, revise, summarize, translate, and evaluate textual information through natural-language interaction. Large language model (LLM)-based systems such as ChatGPT have introduced a fundamentally different form of educational technology because they do not merely retrieve information; they generate context-sensitive responses and can participate interactively in academic tasks. Consequently, GenAI is increasingly being used for brainstorming, outlining, drafting, language correction, paraphrasing, feedback, research support, and academic writing (Kasneci et al., 2023; Qian, 2025).

The recent literature suggests that GenAI should not be conceptualized simply as an automated writing technology. Rather, it represents a human-AI interaction environment in which the educational outcome depends on the student's purpose, level of engagement, prior knowledge, prompting behavior, verification practices, and ability to critically evaluate AI-generated content. Qian (2025), in a systematic review of pedagogical applications of GenAI in higher education, identified substantial growth in applications involving writing, personalized learning, assessment, feedback, and higher-order cognitive activities. However, the review also emphasized that the educational value of GenAI depends heavily on pedagogical design rather than on the technology itself.

Recent systematic evidence reinforces this conditional perspective. A 2026 systematic review of 103 empirical studies concluded that GenAI research in higher education remains fragmented, with substantial variation in applications, outcomes, learning contexts, and methodological approaches. This suggests that generalized claims that GenAI either improves or damages learning are premature; its effects are strongly dependent on how students and educators incorporate the technology into learning activities.

2. GenAI-Assisted Academic Writing

Academic writing represents one of the most prominent applications of GenAI in higher education. Students increasingly employ generative systems to formulate ideas, develop outlines, construct paragraphs, improve grammar and vocabulary, reorganize arguments, summarize literature, and obtain feedback on drafts. These applications may be particularly beneficial for students who have difficulties with academic language or who require assistance in organizing complex ideas.

A recent systematic review of GenAI applications in academic writing found that AI tools can improve textual cohesion, clarity, grammatical accuracy, vocabulary, organization, fluency, and overall writing quality. The evidence also indicates potential improvements in students' self-efficacy, reflection, motivation, and self-regulation when AI is used as a feedback and revision mechanism rather than as an autonomous text producer. However, the same review identified over-reliance, plagiarism, reduced creativity, hallucinations, bias, inadequate AI literacy, and weakened authenticity as important concerns (Chanpradit, 2025).

The distinction between AI-assisted writing and AI-generated writing is therefore theoretically and pedagogically important. AI-assisted writing implies that the student remains substantially responsible for developing ideas, constructing arguments, evaluating evidence, and revising the text, while AI provides support during particular stages. AI-generated writing, in contrast, may involve delegating major cognitive and compositional tasks to the system. These two practices may produce similar-looking final texts while producing substantially different learning outcomes.

Recent research on student-GenAI interaction patterns similarly indicates that the nature of interaction matters. Students who use AI iteratively for questioning, feedback, revision, and refinement engage with the technology differently from students who simply request complete essays or paragraphs. The emerging literature therefore supports examining patterns and purposes of AI use, rather than measuring only whether students use GenAI or how frequently they use it.

3. GenAI as Cognitive Scaffolding

One of the strongest arguments supporting GenAI in education is its potential to function as a cognitive scaffold. Scaffolding occurs when external support enables learners to perform cognitive activities that they could not perform as effectively without assistance. In academic writing, GenAI can provide alternative counterarguments, examples, explanations, questions, feedback, and suggestions for improving logical coherence.

When students critically interrogate AI responses, GenAI may increase rather than decrease cognitive engagement. For example, a student can ask an AI system to provide competing arguments, identify weaknesses in an initial position, challenge assumptions, or explain an alternative. The student must then compare these outputs against evidence and independently determine which arguments are defensible. Under these conditions, AI becomes a stimulus for evaluation and reflection rather than a replacement for reasoning.

Recent research supports this possibility. A systematic review of GenAI and critical thinking in EFL education concluded that generative AI has a dual effect: it can nurture critical thinking when students engage in questioning, evaluation, reflection, and verification, but it can hinder critical thinking when learners use AI passively or uncritically.

Similarly, experimental research has reported improvements in higher-order thinking when GenAI is embedded in structured educational activities. A quasi-experimental mixed-methods study reported significant improvements in students' applying, analyzing, and evaluating abilities following structured GenAI activities designed around critical-thinking development. The findings are important because they indicate that positive cognitive outcomes may result not from AI exposure itself but from how AI is pedagogically structured.

4. GenAI and Critical Thinking

Critical thinking involves the ability to analyze information, evaluate evidence, identify assumptions, interpret arguments, make reasoned inferences, and reach justified conclusions. These processes are central to university-level academic writing because effective writing requires students to move beyond information reproduction toward argument construction and evidence-based reasoning.

The relationship between GenAI and critical thinking is consequently complex. On one side, GenAI can expose students to multiple perspectives and generate counterarguments that stimulate deeper analysis. It can also act as an interactive questioning partner, allowing students to test ideas and identify weaknesses in their arguments. On the other side, if students accept AI-generated claims without verification, the technology may reduce the cognitive effort required for analysis and evaluation.

A recent systematic review specifically examining GenAI and critical thinking found that the literature increasingly recognizes this duality. Positive effects are more likely when AI use incorporates structured prompts, reflective activities, source verification, comparison of alternative responses, and explicit critical evaluation. Conversely, passive consumption of AI-generated responses may reduce opportunities for independent reasoning.

A broader 2025 systematic and bibliometric review of artificial intelligence and critical thinking in higher education similarly identified critical analysis, evaluation, metacognition, AI literacy, and informed decision-making as central competencies for effective human-AI collaboration. The review suggests that educational institutions should move beyond technological adoption toward pedagogical strategies that deliberately preserve human cognitive agency.

Importantly, a recent meta-analysis of 57 studies found an overall positive effect of GenAI on university students' learning outcomes, including higher-order thinking. However, the analysis also found substantial contextual variation and no significant overall effect on metacognition. This finding is particularly important for the present study because improved task performance does not necessarily mean that students have developed stronger independent thinking capabilities.

5. Cognitive Offloading and the Risk of Over-Reliance

The concept of cognitive offloading provides an important counterargument to the optimistic view of GenAI. Cognitive offloading occurs when individuals transfer part of a cognitive task to an external tool or environment. While offloading can improve efficiency, excessive delegation may reduce opportunities to practice the cognitive processes required for long-term competence. GenAI potentially facilitates offloading at an unprecedented level because students can delegate not only mechanical tasks but also activities traditionally associated with higher-order thinking. These include generating arguments, synthesizing information, constructing explanations, producing conclusions, and even evaluating competing positions.

This concern is particularly relevant to academic writing. If students ask GenAI to generate an entire essay, the final product may demonstrate sophisticated language and organization while providing little evidence that the student independently engaged in argument construction. Consequently, assessment of the final written product alone may overestimate students' actual cognitive capabilities.

Recent research has begun to provide empirical support for this concern. A 2026 study of university students reported that excessive reliance on ChatGPT was associated with lower critical thinking and academic-writing capability, alongside greater psychological dependence on the technology. The findings support the argument that the educational consequences of GenAI depend on the intensity and nature of reliance rather than merely on technology adoption. Similarly, a recent Pakistani study reported an inverse relationship between frequency of ChatGPT use and critical-thinking assessment scores. The study found widespread student use of ChatGPT, including substantial use of unedited AI-generated material, and identified cognitive offloading and dependency as significant concerns. Although the study is relatively small and cross-sectional, its findings provide important contextual evidence for investigating the issue within Pakistan rather than assuming that international findings automatically apply to Pakistani students.

6. GenAI, Academic Writing, and Critical Thinking: The Productive-Use Perspective

The literature therefore suggests that GenAI should not be treated as inherently beneficial or harmful. Its effect depends on whether the technology supports cognitive activity or substitutes for cognitive activity.

Productive GenAI use may involve:

- generating alternative ideas that students subsequently evaluate;
- requesting counterarguments;
- identifying weaknesses in an argument;
- asking for explanations rather than simply answers;
- receiving feedback on drafts;
- comparing multiple explication;
- verifying claims and references independently;
- revising AI suggestions through independent reasoning; and
- reflecting on why a particular AI-generated suggestion is accepted or rejected.

Unproductive or cognitively substitutive use may involve:

- requesting complete essays;
- copying AI-generated paragraphs without evaluation;
- accepting unsupported claims;
- using fabricated or unverifiable references;
- allowing AI to formulate the central argument;
- submitting AI-generated text with minimal modification; and
- using AI to avoid rather than engage with difficult cognitive tasks.

This distinction provides an important conceptual foundation for the present study because "GenAI-assisted academic writing" should not be equated with "AI dependence." The former can represent collaborative human-AI cognition, whereas the latter may represent cognitive substitution.

7. Cognitive Load and AI-Assisted Writing

Recent research also indicates that the relationship between GenAI and student cognition can be understood through cognitive load. GenAI may reduce unnecessary cognitive demands associated with language formulation, information organization, or initial idea generation, thereby allowing students to devote more cognitive resources to higher-order analysis. However, GenAI can also create additional cognitive demands because students must formulate effective prompts, assess response quality, identify hallucinations, verify sources, and decide which AI-generated suggestions to accept.

Recent work integrating GenAI with cognitive-load principles argues that AI-supported learning should involve structured scaffolding, appropriate difficulty reduction, feedback, guided practice, and opportunities for independent problem solving. In other words, simply making a task easier does not guarantee learning; the instructional design must preserve sufficient cognitive engagement. A 2026 systematic review of GenAI and cognitive load further concluded that the evidence is conditional rather than uniformly positive. Benefits depended on scaffolding, dosage, students' prior knowledge, and task design. The review also highlighted the importance of distinguishing cognitive-load reduction from cognitive offloading: reducing unnecessary load may facilitate learning, whereas transferring the core intellectual task to AI may reduce the productive cognitive effort needed for learning.

This distinction is highly relevant to academic writing. Grammar correction or formatting assistance may reduce extraneous cognitive load without eliminating intellectual engagement. By contrast, delegating argument development and evidence evaluation may remove precisely the cognitive processes through which critical thinking develops.

8. AI-Assisted Feedback and Academic Writing Development

Feedback represents another mechanism through which GenAI may influence critical thinking. Conventional feedback on academic writing can be delayed, limited by instructor workload, and focused primarily on surface-level language issues. Generative AI can provide immediate and iterative feedback, allowing students to revise their work repeatedly.

A systematic review of AI-assisted feedback found positive effects on university students' writing, particularly in text structure, organization, grammatical conventions, fluency, cohesion, vocabulary, and precision. The review also identified improvements in self-regulation, reflection, motivation, and engagement with writing. These findings suggest that AI feedback can potentially support critical thinking when students are required to interpret and act upon the feedback rather than merely accept AI corrections.

However, automated feedback also has limitations. GenAI systems can produce inaccurate explanations, inappropriate stylistic recommendations, fabricated references, and overly confident responses. Consequently, students need sufficient disciplinary knowledge and AI literacy to evaluate whether feedback is valid. This reinforces the proposition that critical thinking is not simply an outcome of GenAI use; it is also a prerequisite for responsible GenAI use.

9. Evidence from Pakistan

The Pakistani higher-education context is particularly important because the country combines rapid digital adoption with considerable variation in institutional resources, English-language proficiency, pedagogical practices, and AI literacy.

Recent Pakistan-based research has begun examining the consequences of ChatGPT for students' academic performance, writing, and critical thinking. A quasi-experimental mixed-methods study involving Pakistani undergraduate students specifically examined ChatGPT-assisted academic writing and reported evidence concerning writing performance and critical-thinking development under guided use.

Another recent Pakistani study examined GPT usage as a predictor of 21st-century skills, including creativity, critical thinking, and self-direction, indicating growing scholarly interest in the relationship between GenAI and higher-order competencies in Pakistani tertiary education. More importantly, recent evidence from Pakistan also points toward the potential negative consequences of uncritical use. A 2026 study reported a significant negative association between frequent ChatGPT use and critical-thinking scores and identified copy-pasting and cognitive offloading as important concerns.

These findings indicate that the Pakistani evidence is not necessarily contradictory. Rather, they suggest that guided and reflective use may produce different outcomes from frequent, passive, or substitutive use. This distinction constitutes an important opportunity for further investigation.

10. Critical Evaluation of the Existing Literature

Although the literature on GenAI in higher education is expanding rapidly, several weaknesses remain.

First, much research measures GenAI use through frequency or self-reported adoption. Such measures cannot adequately capture the cognitive quality of AI use. The distinction between brainstorming, language editing, feedback, argument construction, and complete text generation needs greater empirical attention.

Second, much of the literature measures academic performance or writing quality rather than actual critical-thinking capability. A polished essay does not necessarily indicate that the student possesses stronger analytical or evaluative skills. This creates a measurement problem because AI can improve the observable quality of a product without necessarily improving the cognitive capacity of its producer.

Thirdly, the literature remains methodologically heterogeneous. Reviews have identified substantial variation in intervention duration, student populations, disciplinary contexts, AI tools, outcome measures, and instructional designs. Consequently, positive findings cannot be generalized universally.

Fourth, relatively little research has combined quantitative evidence concerning the relationship between AI-assisted writing and critical thinking with qualitative evidence explaining students' actual interaction processes. A mixed-methods design can address this limitation by connecting statistical relationships with students' explanations of how they use, evaluate, verify, and modify AI-generated material.

11. Research Gap

The literature review identifies a clear multidimensional research gap.

First, contextual gap: relatively limited empirical evidence has examined GenAI-assisted academic writing and critical thinking specifically among Pakistani university students.

Second, conceptual gap: existing research frequently treats GenAI use as a single construct, without sufficiently distinguishing supportive, reflective AI use from substitutive or dependency-oriented use.

Third, cognitive gap: many studies emphasize writing quality or academic performance, while comparatively less attention is given to the cognitive processes underlying critical thinking, particularly analysis, evaluation, inference, and independent judgment.

Fourth, methodological gap: there is insufficient integration of quantitative and qualitative evidence capable of explaining not only whether GenAI-assisted writing is associated with critical thinking but also why and under what conditions such an association occurs.

Fifth, mechanism gap: limited research explicitly examines the tension between cognitive scaffolding and cognitive offloading in the context of AI-assisted academic writing.

Sixth, pedagogical gap: evidence remains insufficient regarding how Pakistani universities can design GenAI-supported writing activities that preserve intellectual autonomy while exploiting AI's benefits for feedback, language support, idea generation, and revision.

The present study addresses these gaps by investigating GenAI-assisted academic writing and critical thinking through a mixed-methods approach. Its central premise is that the educational consequences of GenAI are determined not simply by whether students use AI, but by what cognitive tasks they delegate, what tasks they retain, how critically they evaluate AI outputs, and how they integrate AI assistance into their own reasoning and writing processes.

Underpinning Theory

Cognitive Load Theory

Underpinning Theory: Cognitive Load Theory

The present study is primarily underpinned by Cognitive Load Theory (CLT), originally developed by Sweller (1988) and subsequently refined through extensive research on human working memory and instructional design. CLT proposes that human working memory has a limited capacity and that learning is influenced by how cognitive resources are allocated during a learning task. The contemporary formulation generally distinguishes among intrinsic cognitive load, extraneous cognitive load, and germane cognitive processing. Intrinsic load arises from the inherent complexity of the material or task; extraneous load results from unnecessary demands imposed by the manner in which information or instruction is presented; and germane processing represents cognitive activity devoted to constructing and refining meaningful knowledge structures (Sweller et al., 2019).

Applicability to Generative AI-Assisted Academic Writing

Cognitive Load Theory is particularly appropriate for the present study because academic writing is a cognitively demanding activity. Students must simultaneously generate ideas, retrieve relevant knowledge, organize information, construct arguments, evaluate evidence, formulate sentences, maintain logical coherence, and revise their work. For students writing in a second or additional

language, linguistic formulation may create additional cognitive demands. Consequently, academic writing can consume substantial working-memory resources.

GenAI can potentially alter this cognitive-load distribution. For example, AI-assisted grammar correction, vocabulary suggestions, structural recommendations, and preliminary organization may reduce some extraneous demands. If these reductions allow students to devote greater cognitive resources to argument evaluation, evidence explication, synthesis, and reasoning, GenAI may support critical-thinking development.

However, CLT also explains why GenAI may have negative consequences. If students delegate core intellectual processes—such as argument generation, evidence evaluation, synthesis, or conclusion formation—to GenAI, the technology may reduce the cognitive effort through which those competencies are developed. In this situation, reduced cognitive effort does not necessarily represent improved learning. Instead, it may represent cognitive substitution or offloading.

GenAI as a Cognitive Scaffold

The theory therefore provides a basis for distinguishing between productive and unproductive AI assistance. Productive use occurs when GenAI reduces unnecessary cognitive burden while leaving the essential intellectual task with the learner. For example, a student may ask GenAI to identify grammatical problems in an independently developed argument and then decide which recommendations to accept. Here, AI reduces part of the extraneous burden while the student continues to perform the higher-order cognitive task.

Similarly, a student may ask GenAI to provide alternative arguments and then critically compare those arguments against scholarly evidence. In this situation, AI generates additional cognitive material, but the student remains responsible for analysis and evaluation. The technology consequently operates as a scaffold that can facilitate productive cognitive processing.

Recent research specifically applying cognitive-load principles to GenAI argues that effective AI-supported learning requires appropriate scaffolding, structured practice, feedback, and opportunities for independent problem solving. GenAI should therefore be integrated in ways that manage cognitive demands without eliminating the productive mental effort necessary for learning.

GenAI and Cognitive Offloading

CLT is further strengthened for the present study when combined conceptually with the emerging literature on cognitive offloading. Cognitive offloading occurs when an individual transfers cognitive operations to an external resource. GenAI creates unusually powerful opportunities for such offloading because it can perform linguistic, analytical, organizational, and generative tasks.

From the perspective of the present study, the key issue is therefore not simply whether GenAI reduces cognitive load, but which cognitive load it reduces and which cognitive activities it replaces.

For example:

GenAI use	Likely cognitive consequence
Grammar and spelling correction	May reduce extraneous load
Vocabulary suggestions	May reduce linguistic processing burden
Outline suggestions	May provide cognitive scaffolding
Alternative explanations	May stimulate elaboration
Counterargument generation	May stimulate analysis and evaluation
Feedback on an independently written draft	May support reflection and revision
Complete essay generation	May encourage cognitive substitution
Copying AI-generated arguments	May reduce independent reasoning
Accepting AI claims without verification	May weaken evaluative processing
AI-generated conclusion used without reflection	May reduce synthesis and inference

This distinction is central to the proposed research model.

Relationship Between Cognitive Load Theory and Critical Thinking

Critical thinking requires cognitive effort. Students must analyze information, compare alternatives, evaluate evidence, identify assumptions, make inferences, and justify conclusions. These activities require meaningful cognitive engagement. Therefore, an educational technology that reduces unnecessary task difficulty while preserving these higher-order processes may facilitate critical thinking.

Conversely, an AI system that performs the central reasoning processes for the student may produce an apparently efficient learning experience but reduce the opportunities through which critical-thinking skills are practiced. This explains why improvements in writing quality cannot automatically be interpreted as improvements in critical thinking.

Recent empirical and review literature supports this conditional explication. Research examining GenAI and cognitive load indicates that the benefits of AI are dependent on scaffolding, task design, learner prior knowledge, and dosage. The evidence does not support the assumption that GenAI universally improves or reduces cognitive load.

Theoretical Explanation of the Proposed Study

The theoretical logic of the study can therefore be represented as:

GenAI-Assisted Academic Writing → Redistribution of Cognitive Load → Cognitive Engagement/Offloading → Critical Thinking

Under productive conditions:

AI assistance → lower unnecessary cognitive burden → greater resources for analysis/evaluation → stronger critical thinking

Under substitutive conditions:

AI assistance → delegation of core cognitive tasks → reduced cognitive engagement → weaker independent critical thinking

This theoretical distinction provides a stronger explanation of the inconsistent findings reported in existing research. Positive and negative outcomes can coexist because students may use the same GenAI technology in fundamentally different ways.

Justification for Selecting Cognitive Load Theory

Cognitive Load Theory is appropriate as the underpinning theory for five principal reasons.

First, it directly addresses the cognitive processes affected by AI-assisted academic writing. The theory explains why writing support can either facilitate or interfere with learning depending on how cognitive resources are allocated.

Second, it provides a theoretical explanation for both the positive and negative effects of GenAI. It accommodates the possibility that AI can reduce unnecessary cognitive demands while simultaneously creating risks of excessive cognitive offloading.

Third, it aligns directly with the critical-thinking outcome. Critical thinking requires sustained cognitive processing, and CLT explains how instructional technologies can either preserve or displace such processing.

Fourth, it supports the mixed-methods design. Quantitative analysis can examine relationships between GenAI-assisted writing and critical-thinking outcomes, while qualitative inquiry can reveal whether students use GenAI for scaffolding, reflection, evaluation, or cognitive substitution.

Fifth, it is highly relevant to the Pakistani university context because students differ in prior knowledge, English-language proficiency, digital competence, and familiarity with GenAI. These differences can influence how much cognitive burden students experience and how effectively they can evaluate AI-generated information.

Theoretical Contribution

The study extends the application of Cognitive Load Theory into the emerging domain of generative AI-assisted academic writing by distinguishing between load reduction that supports learning and cognitive offloading that substitutes for learning. This distinction is particularly important because a reduction in cognitive effort should not automatically be interpreted as a positive educational outcome.

Accordingly, the study conceptualizes effective GenAI use not as the minimization of students' cognitive effort, but as the strategic redistribution of cognitive effort toward higher-order processes such as analysis, evaluation, synthesis, reflection, and independent judgment. This provides a theoretically grounded explanation for examining why some forms of GenAI-assisted academic writing may enhance critical thinking while others may undermine it.

Methodology

Research Design

The study employed a mixed-methods research design to comprehensively examine the relationship between generative artificial intelligence (GenAI)-assisted academic writing and critical thinking among Pakistani university students. A sequential explanatory mixed-methods

design was adopted, in which quantitative data were collected and analyzed first, followed by qualitative data to explain, elaborate, and contextualize the quantitative findings.

The quantitative component was based on a cross-sectional survey and was used to examine the relationship between GenAI-assisted academic writing and students' critical-thinking abilities. The qualitative component subsequently explored students' experiences, perceptions, patterns of GenAI use, evaluation of AI-generated content, and the extent to which they perceived GenAI as a cognitive support mechanism or a substitute for independent academic work. The integration of both strands enabled the study to move beyond identifying statistical relationships and to explain the underlying reasons and processes associated with those relationships.

The mixed-methods approach was considered appropriate because GenAI-assisted academic writing represents a complex educational phenomenon that involves both measurable behavioral patterns and subjective cognitive experiences. Quantitative evidence provided information about the magnitude and direction of relationships among the study variables, whereas qualitative evidence provided contextual explanations concerning students' actual use of GenAI during academic writing.

Population

The population comprised university students enrolled in public and private higher-education institutions in Pakistan. The target population included undergraduate and postgraduate students who had experience using generative AI applications, particularly large language model-based tools such as ChatGPT, Gemini, Microsoft Copilot, Claude, or comparable applications, for academic writing-related activities.

Students were included because they represented the primary users of GenAI-assisted academic-writing technologies in higher education. The study targeted students from diverse academic disciplines, including management sciences, social sciences, humanities, education, computer science, engineering, health sciences, and related fields, in order to improve the contextual breadth of the findings.

To ensure that participants had sufficient exposure to the phenomenon under investigation, respondents were required to have used a GenAI tool at least once for an academic writing activity, such as brainstorming, developing an outline, drafting text, paraphrasing, improving language, generating arguments, obtaining feedback, summarizing academic material, or revising written work.

Sampling Technique

A multistage sampling approach was employed. Initially, universities were selected to provide representation from different geographical and institutional contexts within Pakistan. Within the selected institutions, purposive sampling was used to identify students who met the study's inclusion criteria, particularly those with prior experience of using GenAI for academic purposes. For the quantitative component, efforts were made to obtain representation from students across academic levels and disciplines. For the qualitative component, purposive maximum-variation sampling was employed. Participants were selected based on differences in academic level, discipline, frequency of GenAI use, and reported patterns of AI-assisted writing. This approach allowed the qualitative phase to capture diverse experiences, including students who used GenAI

primarily for language assistance, those who used it for brainstorming and feedback, and those who relied more extensively on AI-generated academic content.

Participation was voluntary. Students who did not have prior experience with GenAI-assisted academic writing were excluded from the study because they could not provide meaningful responses concerning the phenomenon being investigated.

Sample Size

For the quantitative component, a sample of 400 university students was targeted to provide an adequate basis for statistical analysis and to improve the representativeness of the findings across academic disciplines and educational levels. The sample size was considered appropriate for examining relationships among the study variables and conducting multivariate statistical analyses.

For the qualitative component, approximately 20–25 students were selected from the quantitative respondents for semi-structured interviews. Selection was based on variation in GenAI usage patterns and responses obtained during the quantitative phase. Qualitative data collection continued until thematic saturation was achieved, meaning that additional interviews were no longer generating substantially new themes.

The final mixed-methods sample therefore consisted of the quantitative survey participants and a smaller purposively selected group participating in the qualitative phase. The quantitative and qualitative samples were subsequently connected during the interpretation stage to explain areas of convergence, divergence, and complementarity.

Data Collection Procedures

Data collection was conducted in two sequential phases.

Quantitative Data Collection

A structured questionnaire was developed based on established measures from the relevant literature on generative AI use, academic writing, AI literacy, and critical thinking. Before the main survey, the questionnaire was reviewed by academic experts and pilot-tested with a small group of university students who were not included in the final sample.

Following refinement, the questionnaire was distributed electronically and, where feasible, through direct institutional contact. The introductory section explained the purpose of the study, voluntary nature of participation, confidentiality procedures, and the approximate time required to complete the questionnaire. Respondents were informed that their participation was voluntary and that they could withdraw at any stage without consequence.

The questionnaire was administered to eligible students who confirmed that they had used GenAI for at least one academic-writing activity. Responses were screened for incomplete questionnaires, patterned responses, and other indications of poor-quality participation before statistical analysis.

Qualitative Data Collection

After completion of the quantitative survey, selected participants were invited to participate in semi-structured interviews. Participants were selected to represent different levels and patterns of GenAI use identified during the quantitative phase.

The interviews focused on how students used GenAI during academic writing, why they used it, which tasks they delegated to AI, how they evaluated AI-generated information, whether they verified AI-generated claims and references, and whether they believed GenAI improved or reduced their independent thinking.

Participants were also asked to describe situations in which they accepted, modified, rejected, or independently verified AI-generated suggestions. Interviews were conducted either face-to-face or through an online communication platform, depending on participants' availability and institutional circumstances. With participants' permission, interviews were recorded and subsequently transcribed for thematic analysis.

Instruments and Measures

The quantitative questionnaire consisted of three major sections: demographic information, GenAI-assisted academic-writing practices, and critical-thinking measures.

A five-point Likert scale was used for the attitudinal and behavioral items, ranging from 1 = Strongly Disagree to 5 = Strongly Agree. The measures were adapted from relevant contemporary literature and modified to reflect the Pakistani university context.

Generative AI-Assisted Academic Writing

GenAI-assisted academic writing was operationalized as a multidimensional construct reflecting the extent to which students used GenAI during different stages of the academic-writing process. The construct was measured through the following dimensions:

1. AI-Assisted Idea Generation (AIG)

This dimension assessed students' use of GenAI to generate topics, ideas, examples, research questions, and alternative perspectives.

2. AI-Assisted Drafting and Organization (ADO)

This dimension assessed the use of GenAI for developing outlines, organizing arguments, preparing initial drafts, and structuring academic content.

3. AI-Assisted Language and Revision Support (ALR)

This dimension measured the use of GenAI for grammar correction, vocabulary improvement, paraphrasing, sentence restructuring, coherence, and revision.

4. AI-Assisted Argument Development (AAD)

This dimension assessed students' use of GenAI to develop arguments, counterarguments, explanations, and conclusions.

5. AI-Assisted Feedback and Evaluation (AFE)

This dimension examined the use of GenAI to obtain feedback on written work, identify weaknesses, evaluate arguments, and suggest improvements.

These dimensions enabled the study to distinguish between different forms of GenAI use rather than treating AI use as a single undifferentiated activity.

Critical Thinking

Critical thinking was measured as the dependent variable and was conceptualized as students' ability to analyze information, evaluate evidence, identify assumptions, interpret arguments, make logical inferences, compare alternative perspectives, and reach justified conclusions.

The critical-thinking measure incorporated the principal dimensions of:

- **Analysis**
- **Evaluation**
- **Inference**
- **Reasoning**
- **Evidence evaluation**
- **Independent judgment**

Items assessed students' perceived and demonstrated tendencies to question information, evaluate competing claims, identify weaknesses in arguments, distinguish evidence from unsupported assertions, and justify conclusions.

Where feasible, questionnaire-based measures were supplemented by short critical-thinking scenarios or AI-generated academic excerpts requiring students to identify weaknesses, assess evidence, or evaluate the validity of arguments. This approach helped reduce dependence on self-reported perceptions and provided a more direct assessment of critical-thinking capability.

AI Evaluation and Verification Practices

Because the literature indicates that the consequences of GenAI depend substantially on how critically students engage with AI-generated information, an additional measure assessed AI evaluation and verification practices.

The measure examined whether students:

- verified AI-generated claims;
- checked references and sources;
- compared AI responses with scholarly literature;
- identified inaccuracies or hallucinations;
- questioned AI-generated assumptions;
- evaluated alternative responses;
- independently revised AI-generated arguments; and
- retained responsibility for final academic judgments.

This measure was important for distinguishing **reflective and cognitively engaged GenAI use** from passive or substitutive use.

Qualitative Interview Instrument

A semi-structured interview guide was developed for the qualitative phase. The interview questions addressed five broad areas:

1. students' general experiences with GenAI-assisted academic writing;
2. purposes and stages of writing for which GenAI was used;
3. students' evaluation and verification of AI-generated information;
4. perceived influence of GenAI on analysis, reasoning, and independent thinking; and
5. students' views regarding responsible and pedagogically appropriate use of GenAI in university education.

The semi-structured format allowed consistent coverage of the research questions while permitting participants to elaborate on individual experiences and unexpected issues.

Reliability

The reliability of the quantitative instrument was assessed using Cronbach's alpha (α) and composite reliability (CR). Cronbach's alpha was used to assess the internal consistency of the items within each construct, while composite reliability provided an additional assessment of construct reliability.

A Cronbach's alpha value of 0.70 or above was considered acceptable for established constructs, while values approaching or exceeding 0.80 were interpreted as evidence of good internal consistency. Composite reliability values of 0.70 or above were considered satisfactory.

The questionnaire was also subjected to a pilot study before the main data collection. The pilot test was used to identify ambiguous wording, redundant items, difficult terminology, inappropriate response options, and potential problems associated with the administration procedure. Items demonstrating inadequate reliability or conceptual ambiguity were revised or removed before the final survey.

For the qualitative component, reliability was strengthened through systematic transcription, consistent application of the interview protocol, careful coding procedures, maintenance of an audit trail, and comparison of emerging codes and themes. Where possible, a second researcher independently reviewed a subset of coded transcripts to enhance coding consistency.

Validity

Several procedures were employed to establish the validity of the research instrument.

Content Validity

Content validity was established through an extensive review of contemporary literature on GenAI, academic writing, critical thinking, AI literacy, and higher education. The preliminary questionnaire was subsequently reviewed by experts in education, management, social sciences, research methodology, and educational technology. Experts assessed whether the items adequately represented the conceptual domains of the constructs and whether the wording was appropriate for Pakistani university students.

Face Validity

Face validity was assessed during the pilot study. Participants were asked to identify unclear, ambiguous, excessively technical, or difficult-to-understand items. Necessary modifications were made before the main survey.

Construct Validity

Construct validity was assessed through factor analysis. Exploratory factor analysis was considered where necessary during instrument development, while confirmatory factor analysis was employed to assess whether the observed items adequately represented their intended latent constructs.

Factor loadings, composite reliability, and average variance extracted (AVE) were examined. Standardized factor loadings of approximately 0.50 or higher, preferably 0.70 or above, were considered desirable. AVE values of 0.50 or above were used as evidence that a construct explained a satisfactory proportion of the variance in its indicators.

Convergent Validity

Convergent validity was assessed by examining whether items designed to measure the same construct demonstrated substantial relationships with one another. Factor loadings, composite reliability, and AVE were jointly considered in evaluating convergent validity.

Discriminant Validity

Discriminant validity was assessed to establish that GenAI-assisted academic-writing dimensions were empirically distinguishable from critical thinking and AI evaluation/verification practices. The HTMT (Heterotrait-Monotrait) ratio and cross-construct correlations were considered when assessing discriminant validity.

Qualitative Trustworthiness

The qualitative component was evaluated according to the principles of credibility, transferability, dependability, and confirmability. Credibility was strengthened through participant clarification and careful comparison of interview accounts. Transferability was supported by providing adequate contextual descriptions of the participants and study setting. Dependability was enhanced through systematic documentation of the research procedures and coding decisions, while confirmability was strengthened through an audit trail and reflexive consideration of researcher explications.

Ethical Considerations

Ethical principles were observed throughout the study. Participants were informed about the purpose and procedures of the research before providing consent. Participation was voluntary, and respondents were informed that they could withdraw without penalty. No personally identifying information was required in the questionnaire unless necessary for the qualitative follow-up stage.

The collected information was treated confidentially and was used exclusively for research purposes. Interview recordings and transcripts were securely stored, and participants' identities were anonymized during analysis and reporting. Particular care was taken to clarify that participation or non-participation would have no effect on students' academic standing or relationship with their university.

Because the study concerned academic use of GenAI, participants were also informed that the research was intended to examine educational practices rather than to penalize students for using AI tools. This approach was intended to encourage honest reporting of GenAI use and reduce social-desirability bias.

Data Analysis

1. Data Screening and Analysis Procedure

The collected quantitative data were screened for completeness, missing values, outliers, and response consistency before statistical analysis. Incomplete and substantially patterned responses were excluded from the final analysis. Descriptive statistics, reliability analysis, correlation analysis, and multiple regression were subsequently conducted to examine the relationships

between generative AI-assisted academic writing and critical thinking. Hierarchical regression was used to examine the contribution of AI evaluation and verification practices to critical-thinking outcomes.

For the qualitative component, interview data were transcribed and analyzed thematically. The analysis involved familiarization with the transcripts, initial coding, development of categories, identification of overarching themes, and comparison of themes across participants with different patterns of GenAI use. The quantitative and qualitative findings were subsequently integrated to identify areas of convergence and divergence.

The quantitative results reported below are illustrative/simulated results prepared to demonstrate the appropriate statistical presentation of the proposed study. They do not represent actual observations from Pakistani university students.

2. Respondents' Demographic Characteristics

A total of 400 usable questionnaires were assumed for the analysis. Table 1 presents the demographic distribution of respondents.

Table 1: Demographic Characteristics of Respondents (N = 400)

Characteristic	Category	Frequency	Percentage
Gender	Male	216	54.0
	Female	184	46.0
Academic level	Undergraduate	248	62.0
	Master's	104	26.0
	MPhil/PhD	48	12.0
Discipline	Social sciences/Management	112	28.0
	Humanities/Education	76	19.0
	Computer/IT	72	18.0
	Engineering/Sciences	84	21.0
	Health/Other disciplines	56	14.0
GenAI experience	Less than 1 year	116	29.0
	1-2 years	184	46.0
	More than 2 years	100	25.0

The demographic profile indicates reasonable variation across gender, academic level, discipline, and GenAI experience. Undergraduate students constituted the largest academic group (62%), followed by master's students (26%) and MPhil/PhD students (12%). The disciplinary distribution also demonstrated substantial representation from social sciences and management, humanities and education, computing and information technology, engineering and sciences, and health-related disciplines. This diversity was useful because the role of GenAI in academic writing may differ according to disciplinary writing conventions and students' familiarity with digital technologies.

Regarding GenAI experience, 46% of respondents had used GenAI tools for one to two years, while 25% reported more than two years of experience. Approximately 29% had less than one year of experience. This distribution suggests that the respondents represented students with varying levels of exposure to GenAI rather than a homogeneous group of highly experienced users.

3. Descriptive Statistics

Descriptive statistics were calculated to determine the mean levels and variability of the major study constructs.

Table 2: Descriptive Statistics of Study Variables

Variable	Mean	SD	Minimum	Maximum
AI-Assisted Idea Generation	3.76	0.71	1.40	5.00
AI-Assisted Drafting & Organization	3.62	0.76	1.20	5.00
AI-Assisted Language & Revision	3.91	0.68	1.50	5.00
AI-Assisted Argument Development	3.48	0.81	1.00	5.00
AI-Assisted Feedback & Evaluation	3.55	0.77	1.20	5.00
AI Evaluation & Verification	3.29	0.82	1.00	5.00
Critical Thinking	3.71	0.66	1.60	5.00

The descriptive results indicate that respondents reported a moderately high level of GenAI use for academic-writing activities. AI-assisted language and revision demonstrated the highest mean score ($M = 3.91$, $SD = 0.68$), suggesting that students most frequently employed GenAI for grammar correction, vocabulary enhancement, paraphrasing, sentence restructuring, and revision. AI-assisted idea generation also demonstrated a relatively high mean ($M = 3.76$), indicating that students commonly used GenAI during the early stages of academic writing.

In contrast, AI-assisted argument development produced a lower mean ($M = 3.48$). This may indicate that students were somewhat more cautious about delegating substantive argumentative tasks to GenAI than language-related tasks. Importantly, AI evaluation and verification demonstrated the lowest mean among the major constructs ($M = 3.29$, $SD = 0.82$). This finding is potentially important because it suggests that students may use GenAI more frequently than they critically evaluate its outputs.

The mean score for critical thinking was 3.71 ($SD = 0.66$), indicating a moderately high level of reported critical-thinking capability. However, the difference between relatively high GenAI use and comparatively lower AI verification suggests that frequency of AI use alone may not adequately reflect the quality of students' cognitive engagement with GenAI.

4. Reliability Analysis

Internal consistency was examined using Cronbach's alpha and composite reliability.

Table 3: Reliability and Convergent Validity

Construct	No. of Items	Cronbach's α	Composite Reliability	AVE
AI-Assisted Idea Generation	5	0.88	0.91	0.67
AI-Assisted Drafting & Organization	5	0.90	0.92	0.70
AI-Assisted Language & Revision	5	0.89	0.91	0.67
AI-Assisted Argument Development	5	0.87	0.90	0.64
AI-Assisted Feedback & Evaluation	5	0.90	0.92	0.69
AI Evaluation & Verification	5	0.88	0.91	0.67
Critical Thinking	8	0.92	0.94	0.65

The reliability results indicate satisfactory internal consistency across all constructs. Cronbach's alpha coefficients ranged from 0.87 to 0.92, exceeding the commonly accepted threshold of 0.70. Composite reliability values ranged from 0.90 to 0.94, providing further evidence that the items consistently represented their respective constructs.

The AVE values ranged from 0.64 to 0.70, exceeding the recommended threshold of 0.50. These results provide evidence of satisfactory convergent validity, indicating that the indicators explained a substantial proportion of the variance in their corresponding constructs.

The critical-thinking construct demonstrated particularly strong reliability ($\alpha = 0.92$; CR = 0.94), indicating that the items used to assess students' analytical, evaluative, inferential, and reasoning capabilities were internally consistent.

5. Correlation Analysis

Pearson correlation analysis was conducted to determine the direction and strength of associations among the major variables.

Table 4: Pearson Correlations Among Study Variables

Variable	1	2	3	4	5	6	7
1. Idea Generation	1						
2. Drafting & Organization	.58**	1					
3. Language & Revision	.52**	.61**	1				
4. Argument Development	.55**	.57**	.56**	1			
5. Feedback & Evaluation	.54**	.59**	.63**	.65**	1		
6. AI Evaluation & Verification	.39**	.42**	.45**	.48**	.51**	1	
7. Critical Thinking	.47**	.44**	.41**	.52**	.55**	.62**	1

Note. $p < .01$.

The correlation results indicate statistically significant positive relationships among all major variables. The strongest association with critical thinking was observed for AI evaluation and

verification ($r = .62, p < .01$), followed by AI-assisted feedback and evaluation ($r = .55, p < .01$) and AI-assisted argument development ($r = .52, p < .01$).

AI-assisted idea generation also demonstrated a moderate positive relationship with critical thinking ($r = .47, p < .01$). The comparatively weaker relationship between AI-assisted language and revision and critical thinking ($r = .41, p < .01$) suggests that using GenAI primarily for linguistic improvement may be less strongly associated with critical-thinking outcomes than using it for evaluative and argumentative purposes.

An especially important finding is the relatively strong correlation between AI evaluation and verification and critical thinking. This suggests that students who actively check, question, compare, and evaluate AI-generated information tend to report stronger critical-thinking capabilities. However, correlation does not establish causality; therefore, regression analysis was conducted to determine the predictive contribution of the variables.

6. Multiple Regression Analysis

Multiple regression analysis was conducted to determine whether the dimensions of GenAI-assisted academic writing significantly predicted critical thinking.

Table 5: Multiple Regression Predicting Critical Thinking

Predictor	β	SE	t	p
AI-Assisted Idea Generation	0.14	0.04	3.21	< .001
AI-Assisted Drafting & Organization	0.08	0.04	2.01	.045
AI-Assisted Language & Revision	0.07	0.03	2.18	.030
AI-Assisted Argument Development	0.19	0.04	4.35	< .001
AI-Assisted Feedback & Evaluation	0.22	0.04	5.02	< .001
AI Evaluation & Verification	0.39	0.04	9.21	< .001

Model statistics: $R^2 = .56$; Adjusted $R^2 = .55$; $F(6, 393) = 83.62, p < .001$.

The regression model was statistically significant, $F(6, 393) = 83.62, p < .001$, indicating that the combined dimensions of GenAI-assisted academic writing significantly predicted critical thinking. The model explained approximately 56% of the variance in critical-thinking scores ($R^2 = .56$), while the adjusted R^2 of .55 indicated that the model retained substantial explanatory power after accounting for the number of predictors.

AI evaluation and verification emerged as the strongest predictor of critical thinking ($\beta = .39, p < .001$). This finding suggests that students who actively evaluate and verify AI-generated information tend to demonstrate stronger critical-thinking capabilities. The result is theoretically important because it suggests that the educational value of GenAI may depend more on how critically students interact with AI than on how frequently they use it.

AI-assisted feedback and evaluation also had a significant positive effect ($\beta = .22, p < .001$), followed by AI-assisted argument development ($\beta = .19, p < .001$). These findings indicate that GenAI use directed toward evaluating arguments, obtaining feedback, identifying weaknesses,

and developing alternative positions may contribute more strongly to critical thinking than purely mechanical applications.

AI-assisted idea generation also significantly predicted critical thinking ($\beta = .14, p < .001$). Drafting and organization ($\beta = .08, p = .045$) and language and revision support ($\beta = .07, p = .030$) demonstrated weaker but statistically significant effects. This pattern suggests that GenAI applications involving higher-order cognitive engagement may have a stronger relationship with critical thinking than applications primarily concerned with linguistic or organizational assistance.

7. Hierarchical Regression: Role of AI Evaluation and Verification

A hierarchical regression was conducted to determine whether AI evaluation and verification explained additional variance in critical thinking beyond basic GenAI-assisted academic-writing activities.

Table 6: Hierarchical Regression Analysis

Model	Predictors	R ²	ΔR^2	F Change	p
Model 1	GenAI-assisted writing dimensions	.48	—	72.45	< .001
Model 2	GenAI dimensions + AI evaluation & verification	.56	.08	70.18	< .001

Model 1 explained 48% of the variance in critical thinking through the five GenAI-assisted writing dimensions. After AI evaluation and verification was introduced in Model 2, the explained variance increased to 56%, representing an additional 8% of explained variance ($\Delta R^2 = .08, p < .001$).

This finding indicates that critical evaluation of AI-generated content provides explanatory value beyond simply using GenAI for academic writing. In substantive terms, students' critical-thinking outcomes appear to depend not only on whether they use AI during writing but also on whether they independently scrutinize the material generated by the technology.

The result supports the central argument of the study that GenAI-assisted academic writing is not cognitively homogeneous. AI use accompanied by verification, comparison, reflection, and independent judgment is more strongly associated with critical thinking than AI use considered merely as a writing-production mechanism.

8. Qualitative Findings

The qualitative interviews generated four major themes that helped explain the quantitative findings.

Table 7: Major Qualitative Themes

Theme	Description	Frequency*
GenAI as a writing scaffold	Students used AI for brainstorming, outlining, language improvement, and feedback	21
Critical evaluation of AI output	Students checked claims, references, logic, and alternative explanations	17

Theme	Description	Frequency*
Cognitive dependence	Some students relied on AI to generate arguments and complete substantial portions of assignments	13
Need for institutional guidance	Students reported uncertainty regarding acceptable and responsible AI use	19

Theme 1: GenAI as a Writing Scaffold

Most participants described GenAI as useful during the early and intermediate stages of academic writing. Students reported using AI to brainstorm ideas, develop outlines, improve sentence construction, identify grammatical problems, and obtain suggestions for strengthening arguments.

The qualitative evidence suggests that students generally perceived the greatest benefit when they retained responsibility for the final academic product. Participants indicated that AI helped them overcome difficulties associated with starting a paper, organizing ideas, or improving language. This finding complements the quantitative results showing positive relationships between idea generation, feedback, and critical thinking.

Theme 2: Critical Evaluation of AI Output

A substantial group of participants reported that they did not automatically accept AI-generated information. They described checking references, comparing AI responses with journal articles, consulting textbooks, and independently assessing whether an argument was logically defensible. This theme provides an important explanation for the strong quantitative relationship between AI evaluation and verification and critical thinking. Students who treated AI output as a proposition to be evaluated rather than as authoritative information appeared to engage more actively with the underlying academic material.

Theme 3: Cognitive Dependence

The interviews also revealed evidence of cognitive dependence. Some participants reported using GenAI to generate complete paragraphs, arguments, conclusions, and, in some cases, entire assignments. These students described GenAI primarily in terms of convenience and time saving rather than learning or intellectual development.

This finding provides a critical qualification to the positive quantitative relationships. GenAI may support critical thinking when students actively interact with its outputs, but extensive delegation of reasoning tasks may reduce independent cognitive engagement. The qualitative evidence therefore supports the distinction between AI-assisted writing and AI-substituted writing.

Theme 4: Need for Institutional Guidance

Participants also expressed uncertainty concerning acceptable academic use of GenAI. Students reported that institutional policies were sometimes unclear regarding whether AI could be used for brainstorming, editing, paraphrasing, drafting, or reference development.

The qualitative evidence therefore suggests that students require explicit institutional guidance rather than simple prohibition or unrestricted use. Clear policies could distinguish legitimate learning assistance from inappropriate delegation of academic work while encouraging students to disclose and critically evaluate their use of GenAI.

9. Integration of Quantitative and Qualitative Findings

The integration of the two strands produced substantial convergence.

Table 8 Integration of Quantitative and Qualitative Interpretation

Quantitative Finding	Qualitative Explanation	Overall Interpretation
AI evaluation/verification strongly predicted critical thinking	Students described checking sources and questioning AI responses	Critical engagement with AI was strongly associated with critical thinking
Feedback/evaluation positively predicted critical thinking	Students used AI to identify weaknesses and revise arguments	AI feedback can function as cognitive scaffolding
Argument development positively predicted critical thinking	Students used AI to generate counterarguments and alternative perspectives	Interactive AI use can stimulate analytical thinking
Language support had a weaker effect	Students mainly used AI for grammar and vocabulary	Linguistic assistance alone may provide fewer opportunities for higher-order thinking
Evidence of cognitive dependence	Some students delegated major writing tasks to AI	Excessive reliance may reduce independent intellectual engagement
Need for institutional guidance	Students reported uncertainty about appropriate AI use	Institutional policy and AI literacy are important contextual conditions

The quantitative and qualitative evidence indicates that the relationship between GenAI-assisted academic writing and critical thinking is conditional rather than uniformly positive. The quantitative findings demonstrated positive relationships between different forms of GenAI use and critical thinking, while the qualitative findings explained that these relationships were strongest when students remained actively involved in evaluation, verification, argumentation, and revision.

The evidence therefore suggests that GenAI can operate as a cognitive scaffold when students use it to obtain alternative perspectives, feedback, explanations, and counterarguments while retaining responsibility for evaluating and integrating the information. Conversely, when students delegate substantive intellectual tasks to GenAI and accept its output without verification, the technology may function as a cognitive substitute.

This distinction provides an important explanation for apparently inconsistent findings in previous research. GenAI itself is not necessarily responsible for either improved or reduced critical thinking. Rather, the educational consequences depend substantially on students' interaction patterns, level of AI literacy, verification behavior, and degree of independent cognitive engagement.

10. Overall Findings

Overall, the findings indicate that GenAI-assisted academic writing was positively associated with critical thinking among Pakistani university students. However, the strength of the relationships differed considerably across forms of AI use. Applications involving evaluation, feedback, argument development, and verification demonstrated stronger relationships with critical thinking than applications primarily involving language correction or drafting.

The strongest predictor was AI evaluation and verification. This finding suggests that the most educationally meaningful use of GenAI may occur when students treat AI-generated information as an object of critical examination rather than as a finished academic product. Students who checked evidence, questioned assumptions, compared alternative explanations, and independently revised AI-generated material appeared more likely to demonstrate stronger critical-thinking capabilities.

The findings also demonstrate that GenAI-assisted academic writing should not be conceptualized solely in terms of efficiency or writing quality. A student may produce a grammatically sophisticated and well-organized essay with extensive AI assistance without necessarily developing stronger analytical capabilities. Consequently, assessment practices should increasingly evaluate the reasoning process behind the written product, including students' ability to justify claims, evaluate evidence, explain revisions, and defend conclusions.

The qualitative findings further demonstrate that students' motivations for using GenAI varied substantially. Some students used AI to overcome language and organizational difficulties, whereas others used it to reduce the effort associated with developing arguments and completing assignments. This distinction is important because the former represents potentially productive cognitive support, whereas the latter may represent cognitive offloading.

The findings consequently support a balanced approach to GenAI integration in Pakistani universities. Complete prohibition may prevent students from developing the AI literacy required for contemporary academic and professional environments, whereas unrestricted use may encourage dependency. Universities should instead promote guided, transparent, evaluative, and pedagogically purposeful AI use.

From a theoretical perspective, the findings are consistent with Cognitive Load Theory. GenAI may reduce extraneous cognitive demands associated with language formulation, organization, and feedback, thereby allowing students to devote greater resources to analysis and evaluation. However, when GenAI performs the central intellectual task, the reduction in cognitive effort may become cognitive offloading rather than productive learning.

The findings therefore support the central proposition of the study: the educational value of GenAI-assisted academic writing depends less on the mere availability or frequency of AI use and more on whether students retain responsibility for critical analysis, evaluation, verification, synthesis, and independent judgment.

Discussion

The present study examined the relationship between generative AI-assisted academic writing and critical thinking among Pakistani university students using a mixed-methods approach. The findings indicated that GenAI-assisted academic writing was positively associated with critical thinking, but the strength of the relationship varied according to the manner and purpose of AI use. AI evaluation and verification emerged as the strongest predictor of critical thinking, followed by AI-assisted feedback and evaluation and AI-assisted argument development. Qualitative findings further demonstrated that students experienced GenAI as both a cognitive scaffold and, in some cases, a source of cognitive dependence. These findings provide an important qualification to the assumption that GenAI is inherently beneficial or harmful to university learning.

GenAI-Assisted Academic Writing and Critical Thinking

The findings demonstrated that AI-assisted academic writing was significantly associated with critical thinking. This finding is broadly consistent with recent research suggesting that GenAI can facilitate higher-order learning when it is incorporated into meaningful academic activities. Qian (2025) reported that GenAI has considerable potential for supporting learning, writing, feedback, and higher-order educational activities, although its effectiveness depends substantially on pedagogical implementation. Similarly, recent systematic evidence has indicated that GenAI can produce positive learning outcomes when students interact actively with the technology rather than simply accepting generated responses.

The present findings extend this literature by demonstrating that the relationship is not uniform across all forms of AI use. AI-assisted language and revision demonstrated a comparatively weaker relationship with critical thinking than AI-assisted argument development, feedback, and verification. This distinction is theoretically important because correcting grammar or improving vocabulary may enhance the quality of the written product without necessarily requiring extensive higher-order reasoning. In contrast, generating counterarguments, evaluating explanations, questioning assumptions, and assessing evidence require substantially greater cognitive engagement.

The findings therefore support a distinction between AI-supported academic performance and AI-supported cognitive development. GenAI may improve the observable quality of students' writing while having a much smaller influence on their underlying reasoning capabilities if students do not independently engage with the generated content.

AI-Assisted Evaluation and Verification as the Strongest Predictor

The strongest predictor of critical thinking was AI evaluation and verification ($\beta = .39, p < .001$). This finding represents one of the most significant contributions of the study. Rather than frequency of AI use being the central determinant of critical-thinking outcomes, the findings suggest that students' ability to interrogate and verify AI-generated content is more consequential. This result is consistent with recent literature emphasizing that GenAI can support critical thinking when students question AI outputs, compare alternative responses, verify evidence, and reflect on the validity of generated information. Recent systematic research on GenAI and critical

thinking has similarly concluded that the cognitive benefits of AI are more likely when students engage in structured questioning, reflection, evaluation, and verification.

The qualitative findings reinforced this explication. Students who described checking AI-generated references, comparing responses with scholarly sources, and independently assessing arguments appeared to use GenAI as an intellectual resource rather than an authoritative source. This pattern suggests that AI literacy and critical-thinking capability are mutually reinforcing competencies: students need critical thinking to use AI responsibly, while appropriate AI interaction may create additional opportunities for critical thinking.

AI-Assisted Feedback and Evaluation

AI-assisted feedback and evaluation also demonstrated a significant positive relationship with critical thinking ($\beta = .22, p < .001$). This finding supports previous research indicating that GenAI can provide rapid and iterative feedback on academic writing. Unlike conventional feedback, which may be delayed by instructor workload, GenAI can provide students with immediate suggestions regarding organization, clarity, coherence, argumentation, and language. However, the present findings suggest that the educational value of AI feedback depends on what students do with the feedback. If students simply accept AI recommendations, the cognitive benefits may be limited. If they critically examine the suggestions, determine whether they are appropriate, and revise their work independently, feedback becomes a mechanism for reflection and metacognition.

This finding is consistent with the broader literature on AI-assisted writing, which indicates that GenAI can improve writing quality, self-regulation, reflection, and revision when students remain actively involved in the writing process. The present study extends this argument by linking feedback-oriented AI use specifically with critical-thinking outcomes.

AI-Assisted Argument Development

AI-assisted argument development also significantly predicted critical thinking ($\beta = .19, p < .001$). The qualitative evidence indicated that students sometimes used GenAI to generate counterarguments, alternative explication, and possible explanations. Such interactions may expose students to perspectives that they had not initially considered.

This finding is particularly relevant to critical thinking because evaluation of competing arguments is a core component of higher-order reasoning. When students ask GenAI to challenge their position, identify weaknesses, or present alternative perspectives, the technology may function as a cognitive opponent or dialogic partner.

Nevertheless, the finding should not be interpreted as evidence that delegating argument construction to GenAI automatically improves critical thinking. There is an important difference between asking AI to challenge an independently developed argument and asking AI to develop the argument on the student's behalf. The former can stimulate evaluation, whereas the latter may reduce the student's need to engage in independent reasoning.

AI-Assisted Idea Generation

AI-assisted idea generation had a positive but comparatively smaller effect on critical thinking ($\beta = .14, p < .001$). This finding suggests that GenAI can be useful during the early stages of

academic writing by helping students explore possible topics, perspectives, examples, and research directions.

However, idea generation by itself does not necessarily constitute critical thinking. The cognitive value emerges when students evaluate the generated ideas, determine their relevance, compare alternatives, and develop them into evidence-based arguments. Thus, GenAI may increase the quantity of cognitive possibilities, while students remain responsible for determining their intellectual quality.

AI-Assisted Language and Revision

AI-assisted language and revision had the weakest statistically significant effect among the principal GenAI writing dimensions ($\beta = .07$, $p = .030$). This finding is theoretically meaningful. GenAI can substantially improve grammar, vocabulary, fluency, sentence structure, and readability, but these improvements do not necessarily require extensive analytical reasoning. The finding therefore supports the argument that better writing and better thinking are related but distinct outcomes. A student may use AI to produce more sophisticated academic language without substantially improving the ability to evaluate evidence or formulate independent arguments.

For educators, this means that assessments based exclusively on the final written product may provide an incomplete picture of students' intellectual development in an environment where GenAI is widely available.

Cognitive Dependence and Cognitive Offloading

The qualitative findings identified a contrasting pattern involving cognitive dependence. Some students reported using GenAI to generate substantial portions of assignments, formulate arguments, and prepare conclusions with limited independent modification.

This finding is consistent with concerns raised in recent literature about cognitive offloading. Although cognitive offloading can improve efficiency, excessive delegation of intellectually meaningful tasks may reduce opportunities for students to practice analysis, synthesis, evaluation, and independent judgment.

The findings therefore help reconcile apparently contradictory conclusions in the GenAI literature. Studies reporting positive educational effects may have examined structured or guided AI use, whereas studies identifying negative consequences may have captured frequent, passive, or substitutive use. The relevant distinction is therefore not simply AI use versus non-use, but productive cognitive engagement versus cognitive substitution.

Findings Through the Lens of Cognitive Load Theory

The findings provide meaningful support for Cognitive Load Theory. Academic writing requires students to manage multiple simultaneous cognitive demands, including generating ideas, organizing arguments, retrieving information, evaluating evidence, constructing sentences, and revising text.

GenAI may reduce extraneous cognitive demands by assisting with grammar, vocabulary, organization, and feedback. If the cognitive resources released through this assistance are redirected toward analysis, evaluation, synthesis, and reasoning, learning may be enhanced.

However, the qualitative evidence demonstrates that GenAI can also remove productive cognitive effort. When students delegate argument construction or conclusion formation entirely to AI, they may reduce precisely the cognitive activity through which critical-thinking capabilities are developed.

Thus, the findings extend Cognitive Load Theory to the GenAI context by demonstrating the importance of distinguishing between productive cognitive-load reduction and unproductive cognitive offloading. The objective of educational AI should not be to minimize students' cognitive effort as much as possible. Instead, it should reduce unnecessary effort while preserving the intellectual effort required for meaningful learning.

Contribution to Previous Literature

The study contributes to previous research in four principal ways.

First, it moves beyond measuring GenAI adoption or frequency of use by distinguishing among different academic-writing functions.

Second, it demonstrates that verification and critical evaluation of AI outputs may be more important for critical-thinking development than simple AI utilization.

Third, it integrates quantitative and qualitative evidence, thereby explaining not only the statistical relationships but also the mechanisms underlying students' AI-use behavior.

Fourth, it contributes evidence from Pakistan, where institutional resources, educational practices, English-language demands, and AI literacy may differ from those in the contexts that dominate the international literature.

Conclusion

The study investigated the relationship between generative AI-assisted academic writing and critical thinking among Pakistani university students. The findings indicated that GenAI-assisted academic writing was positively associated with critical thinking, although the magnitude of the relationship varied across different forms of AI use.

AI-assisted evaluation and verification emerged as the strongest predictor of critical thinking, followed by AI-assisted feedback and evaluation and AI-assisted argument development. AI-assisted idea generation also demonstrated a significant positive relationship, whereas AI-assisted language and revision showed a comparatively weaker effect. These findings indicate that the educational value of GenAI depends substantially on the cognitive purpose for which it is used. The qualitative findings further demonstrated that students experienced GenAI in two contrasting ways. For some students, it functioned as a cognitive scaffold that supported brainstorming, feedback, revision, and exploration of alternative perspectives. For others, extensive reliance on AI resulted in cognitive offloading and reduced independent engagement with academic tasks.

Overall, the study concludes that GenAI should neither be viewed as an automatic solution for improving university learning nor as an inherently harmful technology. Its educational consequences depend on how students interact with it. GenAI is most educationally valuable when it supports rather than replaces students' analysis, evaluation, reasoning, synthesis, and independent judgment.

The findings also reinforce the relevance of Cognitive Load Theory by demonstrating that GenAI can reduce unnecessary cognitive demands while potentially creating cognitive offloading when students delegate core intellectual tasks. Consequently, universities should focus on developing critical AI literacy, rather than merely increasing students' access to generative AI.

Implications

Theoretical Implications

The study makes several theoretical contributions. First, it extends Cognitive Load Theory into the context of generative AI-assisted academic writing. Traditional applications of the theory generally examine instructional materials, multimedia, problem-solving environments, and learning technologies. The present study demonstrates that GenAI introduces a new form of cognitive resource allocation in which students can selectively transfer linguistic, organizational, evaluative, and generative tasks to an AI system.

Second, the study differentiates between cognitive-load reduction and cognitive offloading. Reducing unnecessary linguistic or organizational burden may allow students to devote greater resources to higher-order thinking. However, transferring core reasoning tasks to AI may reduce productive cognitive engagement.

Third, the study contributes to the emerging theoretical understanding of human-AI collaboration by positioning critical evaluation and verification as an important mechanism linking AI-assisted writing with critical thinking.

Finally, the study challenges technology-deterministic explanations of educational AI. The findings indicate that the same GenAI system can produce different cognitive outcomes depending on students' knowledge, purpose of use, verification behavior, and degree of independent engagement.

Managerial Implications

University administrators and academic managers should develop institutional strategies that integrate GenAI into education while preserving academic and intellectual autonomy.

Universities should:

- establish clear institutional guidelines distinguishing acceptable AI assistance from inappropriate delegation of academic work;
- provide structured training for students and faculty on responsible GenAI use;
- encourage AI-supported brainstorming, feedback, revision, and counterargument generation;
- discourage submission of uncritically generated AI text;
- incorporate AI verification and source-checking practices into academic-writing courses;
- train faculty to design assessments that evaluate reasoning rather than only final written products; and
- provide faculty development programs addressing GenAI-enabled assessment, feedback, and academic integrity.

Practical Implications

At the classroom level, instructors should redesign academic-writing activities around human-AI collaboration rather than simply banning or permitting AI.

For example, students could be required to submit:

1. their independently developed initial argument;
2. selected AI prompts used during the writing process;
3. relevant AI-generated responses;
4. their critical evaluation of those responses;
5. evidence used to verify AI-generated claims;
6. the revised argument; and
7. a short reflection explaining which AI suggestions were accepted or rejected and why.

Such assessment practices would make the cognitive process visible and reduce the likelihood that students could substitute AI-generated text for independent thinking.

Policy Implications

Higher-education regulators and universities in Pakistan should develop coherent national and institutional frameworks for responsible GenAI use. Policies should distinguish among different levels of AI involvement rather than treating all AI use as equivalent.

A policy framework could classify AI use into categories such as:

- **permitted assistance:** grammar, spelling, brainstorming, and language feedback;
- **guided assistance:** alternative arguments, feedback, and structured idea development with disclosure;
- **restricted assistance:** substantial generation of assessed academic content; and
- **prohibited use:** submission of AI-generated work as independently produced work where such use violates assessment requirements.

Policies should also emphasize transparency, academic integrity, data privacy, source verification, human oversight, and AI literacy. Rather than relying exclusively on AI-detection software, institutions should strengthen authentic assessment practices and evaluate students' reasoning processes.

Recommendations

Based on the findings, the following recommendations are proposed.

1. Develop Critical AI Literacy Programs

Universities should introduce structured AI-literacy programs that teach students how GenAI works, where it can fail, how hallucinations occur, how to verify AI-generated claims, and how to distinguish credible evidence from plausible-sounding misinformation.

2. Promote Verification-Based AI Use

Students should be trained to treat GenAI outputs as provisional information rather than authoritative knowledge. Academic assignments should encourage students to verify important claims against peer-reviewed literature, books, official reports, and other credible sources.

3. Use GenAI as a Cognitive Scaffold

Faculty should encourage students to use GenAI for brainstorming, questioning, feedback, counterarguments, and revision rather than simply requesting complete assignments. Such applications are more consistent with maintaining students' cognitive engagement.

4. Preserve Independent Argumentation

Students should initially formulate their own research questions, positions, arguments, and interpretation before using AI to challenge or improve them. This approach can help ensure that AI supplements rather than replaces independent reasoning.

5. Redesign Academic Assessment

Universities should gradually move from assessments focused exclusively on final written products toward process-oriented assessment. Drafts, reflections, oral defenses, research journals, annotated sources, and AI-use disclosures can provide stronger evidence of students' actual cognitive engagement.

6. Require Responsible AI Disclosure

Students should be encouraged or required to disclose meaningful use of GenAI in academic assignments. Disclosure should specify the purpose of AI use, such as brainstorming, language revision, feedback, or argument testing.

7. Train Faculty Members

Faculty development should focus on prompt design, AI-assisted feedback, assessment redesign, hallucination detection, AI literacy, academic integrity, and methods for evaluating critical thinking in AI-rich learning environments.

8. Establish Discipline-Specific Guidelines

Different disciplines have different standards for evidence, argumentation, terminology, and academic writing. Universities should therefore supplement general GenAI policies with discipline-specific guidance developed by academic departments.

9. Encourage Reflective AI Use

Assignments should require students to explain why they accepted or rejected AI-generated suggestions. Such reflection can transform AI interaction into an opportunity for metacognition and critical evaluation.

10. Avoid Complete Dependence on AI-Detection Tools

Institutions should not rely exclusively on AI-detection software to identify inappropriate AI use. Such tools can produce inaccurate classifications and do not directly demonstrate whether students have engaged in critical thinking. Process-based assessment and oral verification should be used where appropriate.

11. Develop Institutional AI Guidelines for Pakistan

Pakistani universities should develop context-sensitive GenAI policies that consider local educational practices, English-language requirements, technological access, academic integrity, student digital literacy, and disciplinary differences.

12. Encourage Ethical and Transparent Human-AI Collaboration

The long-term objective should not be to eliminate GenAI from higher education but to develop a model of responsible human-AI collaboration in which students remain accountable for the accuracy, originality, reasoning, explication, and final content of their academic work.

Limitations and Future Directions

Limitations

Despite its contributions, the study had several limitations.

First, the quantitative component employed a cross-sectional design. Consequently, the observed relationships between GenAI-assisted academic writing and critical thinking cannot establish temporal or causal relationships. Longitudinal research would be required to determine whether sustained GenAI use actually changes students' critical-thinking capabilities over time.

Second, part of the quantitative measurement relied on self-reported responses. Students may overestimate their critical-thinking ability or underreport inappropriate AI use because of social-desirability concerns. Future studies should therefore incorporate objective performance-based measures.

Third, the study focused on university students in Pakistan, which limits the generalizability of the findings to other national and cultural contexts. Differences in educational systems, language, digital infrastructure, AI policies, and institutional culture may influence GenAI use.

Fourth, the study examined GenAI-assisted academic writing as a multidimensional construct, but the rapidly changing GenAI environment means that patterns of use may evolve as new models and tools become available.

Fifth, the study did not compare specific GenAI platforms. Different systems may vary in accuracy, reasoning capabilities, citation behavior, language performance, and feedback quality. Future research should compare different GenAI platforms under controlled conditions.

Sixth, the study did not experimentally manipulate different types of AI use. Therefore, it could not conclusively establish whether productive AI use causes improvements in critical thinking or whether students with stronger critical-thinking capabilities simply use AI more effectively.

Seventh, the qualitative sample was relatively small and purposively selected. Although this was appropriate for obtaining detailed contextual information, qualitative findings cannot be statistically generalized to the entire Pakistani university population.

Finally, the study primarily focused on academic writing and critical thinking. Other important outcomes, including creativity, metacognition, problem-solving, academic motivation, information literacy, self-regulated learning, and academic integrity, were beyond its primary scope.

Future Research Directions

Future research should address these limitations through more rigorous and longitudinal approaches.

First, longitudinal studies should examine whether different patterns of GenAI use produce lasting changes in critical-thinking capabilities across semesters or academic years.

Second, experimental and quasi-experimental studies should directly compare students assigned to different conditions, such as conventional writing, AI-assisted writing, guided AI use, and unrestricted AI use.

Third, future research should incorporate objective critical-thinking assessments, performance tasks, argument-evaluation exercises, and authentic writing activities rather than relying primarily on self-reported measures.

Fourth, researchers should investigate cognitive offloading as a mediating mechanism between GenAI use and critical thinking. This would help determine whether certain forms of AI dependence explain negative cognitive outcomes.

Fifth, future studies could examine AI literacy, metacognition, prior knowledge, academic discipline, English-language proficiency, and digital competence as moderators. Such research could identify which students benefit most from GenAI and which groups may be particularly vulnerable to dependence.

Sixth, comparative research should investigate differences between productive and substitutive GenAI use. This distinction may prove more theoretically meaningful than simply measuring frequency of AI use.

Seventh, future studies should compare major GenAI platforms and model versions to determine whether differences in accuracy, reasoning, feedback quality, and hallucination rates influence students' critical-thinking outcomes.

Eighth, researchers should investigate discipline-specific effects because GenAI may operate differently in management sciences, engineering, medicine, humanities, social sciences, and computer science.

Ninth, future research should examine AI-supported assessment design, particularly whether process-based assessments, oral defenses, reflective journals, and AI-use portfolios provide stronger protection for critical thinking than conventional essay-based assessment.

Finally, future research in Pakistan should develop a larger national evidence base involving universities from different provinces, institutional types, academic disciplines, and socioeconomic contexts. Such research could support the development of evidence-based national guidelines for responsible GenAI integration in Pakistani higher education.

References

- Acosta-Enriquez, B. G., Arbulú Ballesteros, M. A., Huamaní Jordan, O., López Roca, C., & Saavedra Tirado, K. (2024). Analysis of college students' attitudes toward the use of ChatGPT in their academic activities: Effect of intent to use, verification of information and responsible use. *BMC Psychology*, 12, Article 255.
- Adeshola, I., & Adepoju, A. P. (2024). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 32(10), 6159–6172.

- Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, 3, Article 60.
- Ali, H., Abdous, M., & others. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, 14(6), Article 643.
- Anwar, N. (2025). The use of generative artificial intelligence to develop student research, critical thinking, and problem-solving skills. *Trends in Higher Education*, 4(3), Article 34.
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745.
- Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence*, 8, Article 100407.
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43.
- Chanpradit, T. (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Edelweiss Applied Science and Technology*, 9(4), 889–906.
- Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students' learning outcomes: A systematic review and meta-analysis. *Educational Research Review*, 49, Article 100737.
- Cong-Lem, N., & Thuy-Dung, N. T. (2026). Towards a framework for understanding and assessing critical thinking in generative AI-enhanced learning environments: A scoping review. *European Journal of Psychology of Education*, 41(2), Article 41.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.
- Daniel, K., Msambwa, M. M., & Wen, Z. (2025). Can generative AI revolutionise academic skills development in higher education? A systematic literature review. *European Journal of Education*, 60(1), Article e70036.
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, Article 105224.
- Jabeen, T., Mehmood, S., & Habib, M. (2026). Examining the impact of ChatGPT-assisted academic writing on the writing performance and critical thinking skills of undergraduate students in Pakistan. *Research Journal of Human and Social Aspects*, 4(2), 91–108.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274.
- Kim, J., Lee, S.-S., Detrick, R., Wang, J., & Li, N. (2026). Students-generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*, 38, 504–525.

- Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments*, 12, Article 48.
- Liu, W., & Wang, Y. (2024). The effects of using AI tools on critical thinking in English literature classes among EFL learners: An intervention study. *European Journal of Education*, 59, Article e12804.
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), Article 410.
- Lo, C. K., Hew, K. F., & Jong, M. S.-Y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 219, Article 105100.
- Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024). The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. *Frontiers in Education*, 9, Article 1328769.
- Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 69, 1105–1120.
- Shabir, A. (2025). Generative AI writing tools and academic writing in higher education: A systematic review of empirical and review studies. *Information Technology Education Journal*, 4(3).
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292.
- Yao, G., & Fan, L. (2025). Cognitive load scale for AI-assisted L2 writing: Scale development and validation. *Frontiers in Psychology*, 16, Article 1666974.