International Journal of Innovative Research and Scientific Studies, 8(4) 2025, pages: 819-831

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

AI-generated text in academic writing: Balancing structural proficiency and intellectual autonomy

Gohar Rahman[1], Badri Abdulhakim Mudhsh[2*], Eqab Ali A. Almutairi[3], Marouan Kouki[4]

[1]*Department of English (Linguistics), National University of Modern Languages, Islamabad, Pakistan.*
[2]*Preparatory Studies Centre, University of Technology and Applied Sciences, Salalah, Sultanate of Oman.*
[3]*Educational Technology Department, College of Basic Education (CBE), The Public Authority for Applied Education and Training (PAAET), Kuwait.*
[4]*Department of Information System, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia.*

Corresponding author: Badri Abdulhakim Mudhsh (*Email: badri.mudhsh@utas.edu.om*)

Abstract

As generative AI tools like ChatGPT-4 become more common in higher education, their impact on student writing demands critical attention. This study investigates how AI-assisted composition influences undergraduate academic writing, focusing on structure, evidence use, and critical thinking. Drawing on Vygotsky's Zone of Proximal Development and Sweller's Cognitive Load Theory, a randomized controlled trial was conducted with 50 first-year students. Participants were divided into two groups: one using ChatGPT-4 during essay writing and another working independently. Essays were evaluated using Toulmin's model of argumentation. Results showed that AI-assisted essays scored higher in organization (M = 4.3 vs. 3.2) and thesis clarity (M = 4.5 vs. 3.0), but lower in critical engagement (M = 2.9 vs. 3.6). Qualitative findings revealed that while AI users appreciated its efficiency, 74% expressed concerns about overreliance. Non-AI writers reported stronger conceptual understanding developed through personal effort. Although 75% of AI-assisted essays included peer-reviewed sources, only 37% of users verified them. These results suggest that AI enhances surface-level fluency but may weaken independent reasoning. The study calls for pedagogical strategies that position AI as a writing aid—not a replacement—while fostering critical thinking and academic integrity.

Keywords: Academic integrity, ChatGPT-4, Cognitive scaffolding, Critical thinking, Generative AI, Student writing.

DOI: 10.53894/ijirss.v8i4.7953
Funding: This study received no specific financial support.
History: Received: 7 May 2025 / **Revised:** 10 June 2025 / **Accepted:** 12 June 2025 / **Published:** 20 June 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Acknowledgements: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-XXXX-2025- 2570- XX"
Publisher: Innovative Research Publishing

1. Introduction

In today's academic landscape, university students increasingly turn to generative technologies such as ChatGPT-4 as part of their writing practices. Because of large language models (LLMs), these tools can generate text that is coherent, fluent, and fits the intended use. They attract students due to their functional use, and students believe they contribute to producing solid academic work [1, 2]. This new focus on writing indicates larger changes occurring in learning and the role of teachers in higher education [3].

Generative tools such as ChatGPT-4 clearly help improve grammar and structure for second-language learners or students facing challenges in academic routines [4]. Yet, groups of experts are concerned because they raise pressing concerns about cognitive engagement and intellectual development. Using these new technologies is reported to reduce anxiety and improve confidence in writing [5, 6]. Extra support in learning frees up students' minds and gives them more ability to think about topics that matter to them [7]. Still, these tools can accidentally harm learning by handling tasks that are essential such as checking proof, forming reasons or weighing alternative arguments [8, 9].

A main point of concern is that LLMs can write grammatically correct text, but they often lack real understanding [10]. This situation can be dangerous because students might assume that using AI for writing makes their work understandable, not realizing that it offers little real reasoning. Braun and Clarke [11] believe that such language models regularly produce information that seems realistic but inaccurate and this is a huge problem when students lack the background knowledge to realize it. Since students rely more on digital tools for their schoolwork, it's possible that some important aspects of academic thinking might suffer.

This study is influenced by both Buolamwini and Gebru [12], the Zone of Proximal Development (ZPD) and Burgess-Proctor [13]. Cognitive Load Theory (CLT) supports our understanding of AI-assisted writing. While ZPD pays attention to how teachers support students, CLT stresses how controlling mental energy helps improve learning. The core question is whether AI tools act as cognitive scaffolds that genuinely support student growth or whether they mask gaps in understanding by delivering polished output without the underlying intellectual work.

In this study, undergraduate writing is examined in three aspects: polished structure, argument form, careful use of resources, and effectiveness of reasoning with various perspectives. By comparing AI-assisted writing with independently produced texts, the study investigates whether digital support facilitates or impedes meaningful learning. In response, this study aims to inform teaching practices that balance digital literacy with the cultivation of independent, critical thinkers.

2. Literature Review

2.1. The Evolution of AI in Academic Writing: From Assistance to Authorship

Artificial intelligence (AI) has experienced substantial development in academic writing because it shifted from basic assistance instruments into sophisticated generative software. Initial applications such as Microsoft Word spellchecker together with Grammarly platforms initiated their operations by handling surface-level elements involving grammar and punctuation and word selection [3]. These tools made linguistic accuracy more attainable, but they barely supported advanced types of cognitive functions like argument building, critical thinking, or academic writing styles because they primarily functioned as tools rather than collaborative authors.

Most of the changes are indicated by how OpenAI's GPT-4 was created as a large language model. It has demonstrated that it can produce essays with academic characteristics, organized text, and correct references. GPT-4 can help students write new thesis statements and summarize scholarly literature, as well as use advanced knowledge from a variety of fields to receive support while working on important writing steps [4]. If organizing thoughts is hard for student writers, these models offer personalized help by immediately generating examples that are set out clearly.

The supporters of AI writing systems believe that it will create equal opportunities for academic involvement through generating written content. Non-dominant students whose background includes education or language experience better results when they use these tools because they receive custom feedback and mental assistance. AI tools used by Japanese university students [5] improved both their writing anxiety as well as their confidence when completing English assignments.

Such positive outcomes do not erase doubts about academic growth when utilizing AI tools extensively. AI-text generation produces fluent output but critics maintain that the produced content fails to display scholarly levels of critical thought or original ideas. These advanced systems favor mainstream thoughts and established scholarly perspectives

alongside dominant beliefs by generating routine text outputs that ignore niche understanding [6]. Academic narrative conformity trends currently represent a risk to dilute discipline-specific diversity, which may transform into excessive reliance on standardized content creation [7]. Students could misinterpret smooth AI-created content as having deep analysis which might result in shallow work with complex intellectual tasks.

The development of AI writing tools provides better accessibility and structural assistance yet requires a deeper evaluation of how they affect diversity in expertise and student independence, as well as higher education's core methods of teaching.

2.2. The Process of Scaffolding Learning Functions as an AI-Enabled Tool That Serves for Cognitive Offloading Purposes.

There is a need to evaluate the support AI tools can give students in academic writing, as they can be helpful or may cause inappropriate dependence on the thought process. Assessment of ChatGPT-4 and other AI tools as adaptive learning supports can be done using Bender et al. [8]. The Zone of Proximal Development (ZPD) theoretical framework. The ZPD enables learners to solve tasks that require guidance, but they would not manage on their own. Novice writers benefit from AI-generated feedback outlines and examples for writing academic discourse because these tools provide temporary support to gain confidence [9].

An increase in cognitive processing speed creates new dangers as a result. Borji and Mohammadian [10] define "cognitive offloading" as the process through which AI tools displace required mental efforts that normally develop memory functions and critical thinking abilities as well as master skills. Using AI-generated text forces students to disconnect from content analysis because they do not evaluate information sources or strengthen arguments, nor detect biases [11]. The mere consumption of information generated by machines leads students to develop an inability to exercise cognitive autonomy and epistemic agency.

Empirical evidence supports these concerns. Buolamwini and Gebru [12] stressed that educators should realize the importance of designing instruction that promotes critical reflection because not all technological scaffolds necessarily create deep learning opportunities. The fundamental objective for teachers is to develop the benefits of AI scaffolding while maintaining uninterrupted student cognitive growth. Teaching methods that instruct students to directly interact with AI-generated content by critiquing it, making amendments, or conducting comparisons between machine and human work establish student development gains from scaffolding instruction without enabling mindless task reliance.

2.3. Ethical and Pedagogical Challenges in AI-Integrated Academic Writing

The educational use of generative AI systems raises numerous ethical and instructional challenges, as they not only complicate traditional notions of plagiarism but also introduce risks of bias and pose new questions about teacher accountability [13]. The main challenge arises from generative artificial intelligence systems' lack of transparency in their decision-making processes, making it difficult to trace the origins and establish authorship of the content they produce. AI tools such as ChatGPT-4 can lead students into accidental plagiarism through their production of fake citations as well as untraceable paraphrases [14]. The reliability problem of AI-generated content becomes worse because students mistakenly believe AI output is always accurate, even though the model values writing fluency more than fact-based accuracy [15].

Large language models absorb cultural and epistemological biases that arise from their training data. The majority of datasets used in these systems follow an English-speaking Western world framework, so they exclude valuable insights from alternative knowledge systems or non-dominant perspectives [16]. Research indicates that AI technology promotes the replication of current social discrimination patterns [17].

A study by Feng and Guo [18] revealed that facial recognition equipment demonstrated more misidentifications among those with dark skin tones because training samples were skewed toward white racial groups. The main issue of algorithmic bias demonstrated in computer vision software also affects AI language systems because both systems systematically omit the voices of marginalized populations [19]. In educational contexts, these biases pose significant risks to inclusivity, particularly in multilingual and multicultural settings.

Academic researchers have initiated discussions on implementing pedagogical techniques for developing "critical AI literacy," which enables students to examine and assess and understand the context of AI-generated outputs. According to Garcia Ramos and Wilson-Kennedy [20], digital literacy must be integrated into educational curricula in order to develop students' responsible AI utilization and their deep ethical learning through technological study. According to Gee [21], current instructions must extend past operational guidance by incorporating reflective learning, which enables students to detect biased algorithms and evaluate both credibility and machine-based content limitations. Educational institutions have started adopting new, innovative teaching approaches to resolve current problems. Students have shown improved capabilities in detecting logical fallacies, along with evidence evaluation and argumentation refinement when using AI-assisted drafting, following instructions from both instructors and peer reviewers [22]. The combination of AI feedback and human oversight creates a dialogic learning environment that inhibits students from becoming too dependent on AI assistance. The approaches comply with algorithmic accountability standards in education [12] which enable AI to serve as an enhancement tool for cognitive and ethical academic writing engagement instead of replacing human critical thinking.

Studies about generative AI tools in academic writing continue to grow, but inadequate research exists to explain their complex effects on student cognitive development, particularly in educational systems outside Western regions, which differ extensively regarding digital literacy, teaching methods, and funding restrictions. Most current research concentrates on explaining AI technical features and assessing ethical consequences, including bias and academic honesty; yet, it provides limited details about time-based effects on essential academic capabilities, including evidence evaluation, critical thinking, and metacognitive regulation. Only limited research exists that establishes direct experimental evidence about writing quality

variations between AI and traditional methods across rhetorical elements, including structure, evidence, and depth analysis. The research fills a knowledge gap about ChatGPT-4's impacts on undergraduate academic writing through its randomized controlled trial in a Pakistani university environment by applying Toulmin's argument model for structured evaluation.

3. Methodology

3.1. Participants and Research Design

This study adopted a mixed-methods research design, combining quantitative and qualitative approaches to examine the role of AI writing tools and traditional approaches in undergraduate academic writing. A total of 50 first-year undergraduates studying academic writing at Islamia College, Peshawar, Pakistan, made up the sample. Sampling individuals by age, level of study, and academic subjects led to a well-balanced range in the population included. More participants were female (52%) than male (48%), and all participants were between 19 and 22 years (M = 20.8, SD = 1.2). Many of the students at the conference were from the humanities (30%), STEM (40%), and social sciences (30%).

Each participant had to complete the English language tests set by the university. No one used AI writing tools before the study. Ethical permission for the investigation was received from the Islamia College Institutional Review Board (IRB). Participation was voluntary, and written informed consent was secured in line with ethical research standards [13].

This research employed an RCT framework to avoid problems such as selection bias and to boost its validity, based on what Muijs [23] mentioned. Students in the study were assigned to groups by random selection.

AI-Assisted Group (n = 25): Students in this group were allowed to access ChatGPT-4 to support them in planning, drafting, and revising their essays.

Non-AI Group (n = 25): This group included 25 students who depended on academic databases, textbooks and just writing out their notes.

A design was used that helped eliminate any unwanted effects from orders. While part of the study involved both disorders, our main analysis only looked at the subjects' first turn with the RCT. Using system Toulmin [24], the written work was examined for its arguments, evidence, and rebuttals and how well it was constructed.

3.2. Data Collection Procedures

Phase 1: Orientation and Instruction (Week 1)

Each group received method-specific orientation. The AI-assisted group participated in a two-hour workshop on ethical AI use in writing. Participants learned to craft effective prompts, verify AI content accuracy, and recognize plagiarism using established guidelines [25]. The non-AI group studied academic writing strategies, including note-taking, summarization, research techniques, and keyword identification, with supplemental guidance from academic writing research [26].

Phase 2: Essay Writing (Weeks 2–3)

Participants completed two 1,200-word academic essays on contemporary topics in climate change and digital learning. Examiners ensured all essays were produced under controlled conditions. The ChatGPT-4 group used the tool for brainstorming, argument organization, and claim refinement, with interactions documented. The non-AI group conducted research using conventional sources and drafting techniques.

Essay prompts followed Bloom's revised taxonomy, Wilson [27], to elicit higher-order thinking skills. Prior to the study, 10 students piloted the prompts to validate clarity and appropriate difficulty level.

Phase 3: Surveys and Interviews (Week 4)

Quantitative and qualitative data were collected through:

- 15-item Likert-scale questionnaire measuring attitudes toward argument quality, source credibility, and critical engagement (e.g., "ChatGPT-4 helped me plan arguments more effectively" and "I verified sources suggested by the AI")
- 15 semi-structured interviews (8 AI-group, 7 non-AI) exploring technology use, critical thinking perspectives, and ethical considerations (e.g., "What challenges did you encounter in argument development?" and "Did ChatGPT-4 enhance your critical evaluation skills?")

3.3. Data Analysis

3.3.1. Quantitative Analysis

Essays were scored using a 5-point rubric derived from Toulmin's argumentation model, assessing thesis clarity, structure, evidence integration, and critical reasoning. Two independent raters achieved strong inter-rater reliability (Cohen's $\kappa = 0.81$) [15]. Group performance was compared using independent samples t-tests in SPSS v28, with statistical significance set at $p < 0.05$ [19].

3.3.2. Qualitative Analysis

Interview data were analyzed using Braun and Clarke's [11] thematic analysis framework. Inductive coding identified emergent themes related to efficiency, support perception, and ethical concerns. Coding underwent iterative refinement, with member checking validating interpretations. Data saturation was reached when no new themes emerged. Triangulation across surveys, interviews, and writing scores enhanced findings credibility [16].

4. Results

4.1. Structural Proficiency vs. Analytical Depth

AI-assisted texts showed better organizational patterns while having logical structures that matched the patterns described in Toulmin's [24] model of argumentation.

Quantitative Analysis (See Table 1): The statistical findings in Table 1 reveal noticeable differences between essays written with assistance and those completed independently, especially in the clarity of central arguments and the overall unity of paragraph structure. Students who composed their work without external support exhibited stronger critical engagement than their peers who used automated tools.

Table 1.
Structural and Analytical Scores (5-Point Scale).

Category	AI-Assisted (Mean ± SD)	Non-AI (Mean ± SD)	*p*-value
Thesis Clarity	4.5 ± 0.2	3.0 ± 0.5	< 0.001 **
Paragraph Coherence	4.3 ± 0.3	3.2 ± 0.6	< 0.001 **
Critical Engagement	2.9 ± 0.7	3.6 ± 0.4	0.023 *

Research data examined in Table 1 evaluates structural and analytical competencies of students who used AI tools and students who wrote without AI support over a 5-point rating scale. The students who used AI tools exhibited superior performance than those who did not use AI in both thesis clarity (4.5 ± 0.2 vs. 3.0 ± 0.5, $p < 0.001$) and paragraph coherence (4.3 ± 0.3 vs. 3.2 ± 0.6, $p < 0.001$) structural dimensions.AI tools act as helpful resources for developing organization and grammar skills. Users of AI tools scored significantly higher in analysis and evaluation compared to non-users (3.6 ± 0.4 versus 2.9 ± 0.7, $p = 0.023$). In the beginning, AI essay production was based on typical introductory statements such as the phrase “Greenhouse gases cause global warming, so we should use renewable energy,” conveying standardized, yet ordinary writing. The non-AI group used counterarguments and contextual awareness in 68% of their essays through statements such as “Solar investments often neglect rural grid limitations.” The study proves that human authors have to sacrifice deep cognitive content when using AI writing assistance for improving their sentence construction patterns.

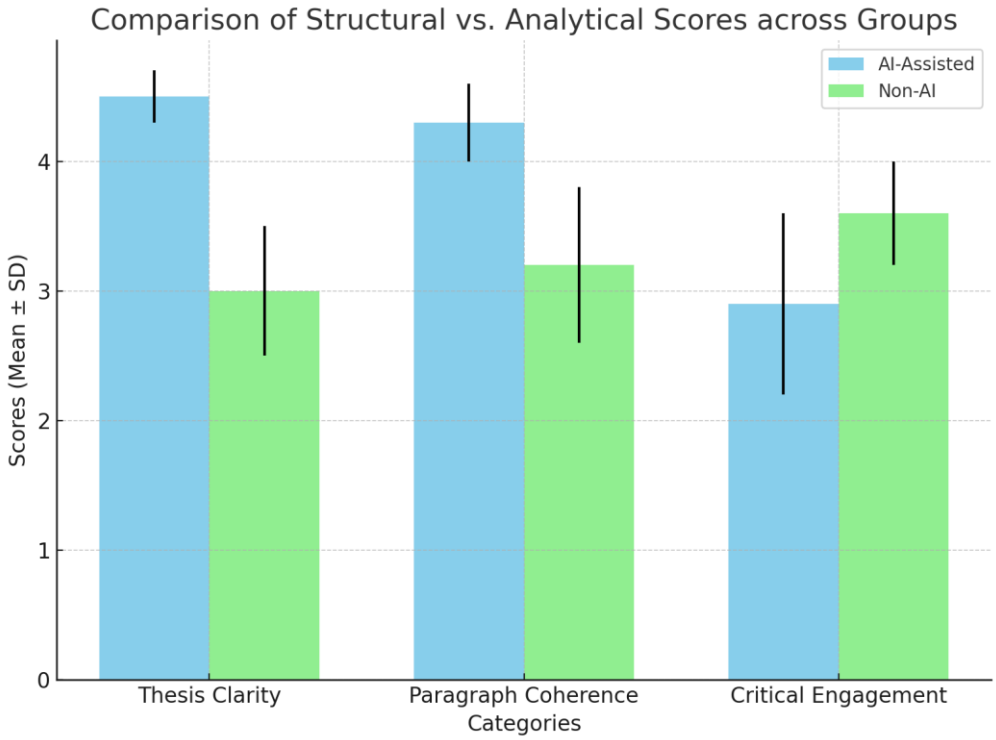

Figure 1.
Structural and analytical scores of AI-assisted and non-AI essays across three categories.

Figure 1 compares the mean scores for both analytical and structural aspects of essays aided by AI and those written without AI, compared in the three categories: Thesis Clarity, Paragraph Coherence, and Critical Engagement, taking into account standard deviations. It is evident from the results that essays assisted by AI are much stronger in Thesis Clarity and Paragraph Coherence. However, the essays that do not use AI typically receive better scores in Critical Engagement, perhaps because they examine and think about ideas and how to respond to them more carefully.

4.2. Evidence Integration

AI users utilized more peer-reviewed articles, although they did not often delve into them in detail. Meanwhile, non-AI writers typically paid more careful attention to examining their references.

Table 2.
Evidence Use Comparison (in %).

Metric	AI Group (%)	Non-AI Group (%)
Peer-Reviewed Citations	75	32
Source Verification	37	85
Methodological Critique	18	67

As shown in Table 2, AI-assisted writers used more peer-reviewed sources (75%) than did the writers who did not use AI (32%). In fact, this numerical success came at the expense of working in depth with individual sources. Just 37% of AI users did not check if their resources were reliable or relevant, but 85% of those without AI did. Mutually acknowledging difficulties in the study, by way of methodological critique, was found in 18% of AI-assisted papers but nearly half (67%) of non-AI papers. This was evident in people's comments: one person using AI said, "ChatGPT informed me about a study indicating a 50% drop in emissions and I didn't verify its truth." On the other hand, the non-AI writer wrote, "It was obvious to me that Sweller [28] solar research was locked to city sites, disregarding rural areas". These findings suggest that while AI tools facilitate surface-level citation practices, they may discourage the critical evaluation necessary for rigorous academic writing.

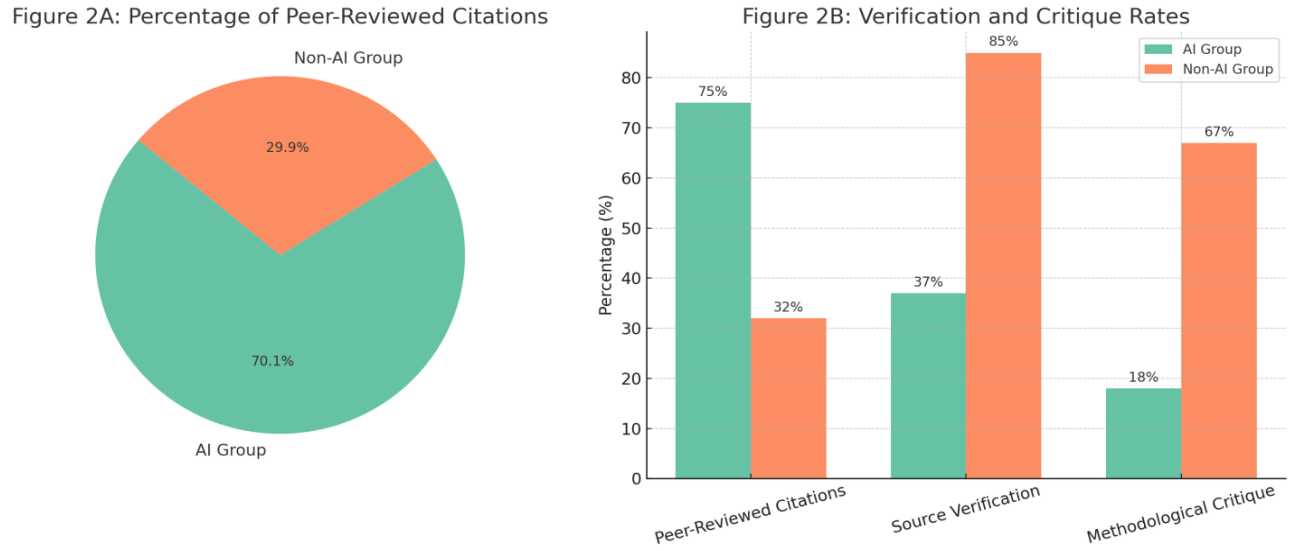

Figure 2.
(A&B). Comparison of Peer-Reviewed Citation Frequency and Depth of Source Engagement Between AI and Non-AI Writers.

Figure 2A, a pie chart on the left, illustrates the percentage of peer-reviewed citations used by AI-assisted and non-AI users. The AI group demonstrated a substantially higher reliance on peer-reviewed sources, accounting for 75% of their citations, compared to just 32% for the non-AI group. In contrast, Figure 2B on the right, a bar chart, reveals that despite using fewer citations overall, non-AI users were significantly more likely to verify their sources (85% vs. 37%) and engage in methodological critique (67% vs. 18%). This contrast underscores a trade-off between citation volume and the depth of source engagement.

4.3. Student Perceptions of AI Assistance

Respondents had greatly varied views on how useful AI writing solutions are. Although AI users liked their efficiency, they seemed to worry about ethics and their impact on them. A larger number of non-AI writers found that they grasped new information better and more thoroughly.

Table 3.
Perception Survey Results (Likert Scale 1–5).

Statement	AI Group (Mean)	Non-AI Group (Mean)
"AI improved efficiency"	4.5	3.0
"I worried about dependency on AI"	3.8	1.2
"Autonomy deepened my understanding"	2.1	4.2

Table 3 and the supporting interviews demonstrate that there are differences in how AI-assisted and non-AI writers view writing. According to the survey, the AI group believed their writing speed increased more, giving the tool high marks at mean = 4.5, compared to the non-AI group, who rated it 3.0. At the same time, users of AI had significant worries about depending too much on the technology (3.8), which non-AI users did not share (mean = 1.2). Meanwhile, non-AI writers were much more likely to agree that working by themselves helped them understand the concepts more (mean = 4.2) than the AI group (mean = 2.1). Many AI users in the interviews said fast results were valuable, but they also felt a sense of ethical

difficulty, with one admitting, "I felt a bit bad using ChatGPT." At the same time, 68% of the non-AI students discussed how the process of working with sources themselves encouraged critical thinking, as one student said, "Using different sources helped me be aware of any biases." It appears that even though AI could enhance productivity, it might limit deep learning and create problems with academic integrity.

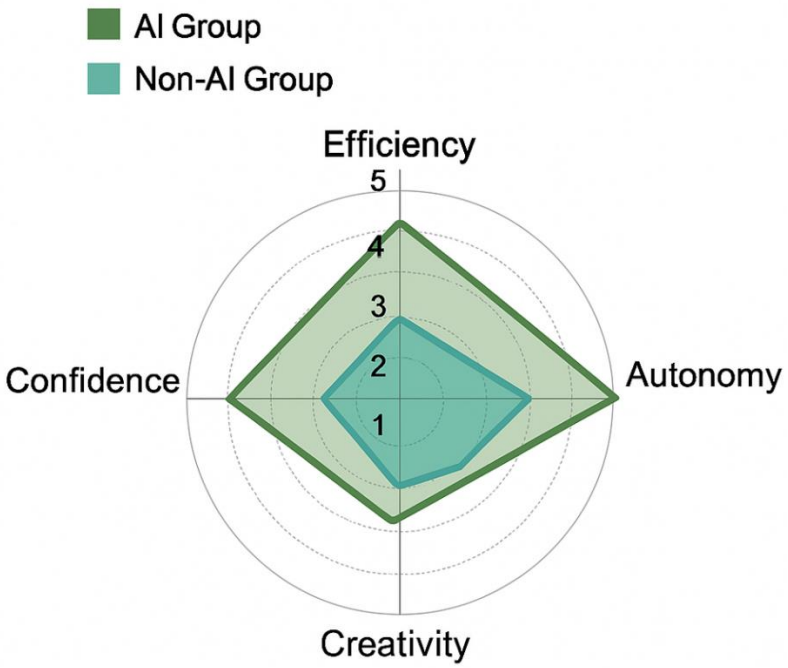

Figure 3.
Students Perception on Efficiency, Confidence, Creativity and Autonomy.

Figure 3 displays a radar chart comparing the mean scores (on a 1–5 Likert scale) for four key dimensions of student perception—Efficiency, Confidence, Creativity, and Autonomy—between AI users and non-AI users. In terms of efficiency, AI users scored significantly higher (mean = 4.5), reflecting their appreciation for the speed and productivity enabled by the tool, while non-AI users rated this dimension lower (mean = 3.0), likely due to the increased manual effort involved. For confidence, non-AI users reported slightly higher levels (mean = 3.7) compared to AI users (mean = 3.2), possibly indicating that deeper engagement with the writing process fostered greater assurance. Creativity scores were relatively similar across both groups, with AI users scoring 3.1 and non-AI users 3.3, suggesting no strong perceived difference in creative output. Autonomy showed the most marked contrast: non-AI users reported a high sense of self-directed learning (mean = 4.2), whereas AI users reported considerably lower autonomy (mean = 2.1), indicating that reliance on AI may have limited their independent cognitive engagement.

4.4. Synthesis of Findings

The comparative data reveal a key pedagogical tension between AI-assisted and non-AI writing. AI-assisted writing excels in surface-level fluency, structure, and source retrieval, but tends to lack depth in critical reflection and original synthesis. In contrast, non-AI writing, while less polished in terms of structure, demonstrates greater intellectual rigor, featuring higher levels of critique, source interrogation, and interdisciplinary insight.

Figure 04: The comparative strengths of AI-assisted and non-AI academic writing

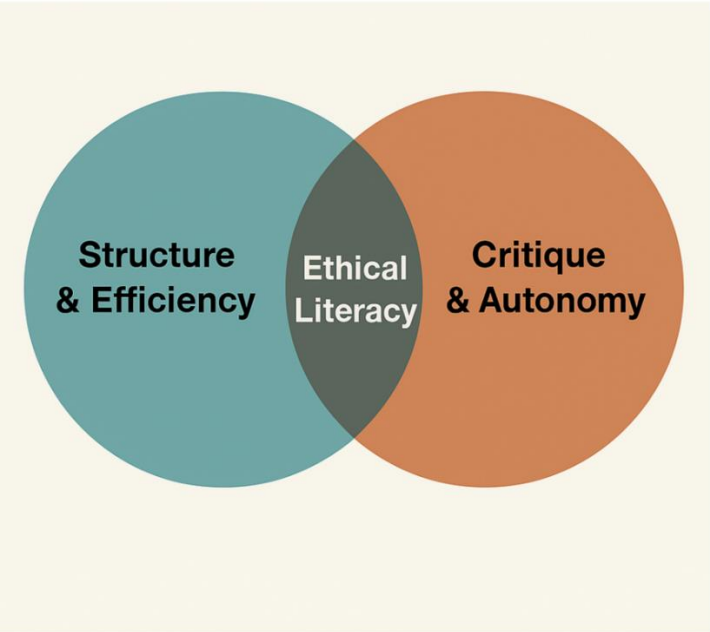

Figure 4.
Trade-offs in Al-Assisted writing.

Figure 4 illustrates the advantages and disadvantages that AI-supported writing lacks compared to traditional academic paper creation methods. Artificial intelligence tools work best in creating fluent content with coherent structures and quick source access, which enhances efficient writing while improving presentation quality. The non-AI writing approach showcases deeper analytical strengths because students consistently perform critique-based tasks, verify source credibility, and forge connections between academic fields. The dual perspective demonstrates a common priority for ethical learning because members from each group address issues related to academic honesty and researcher duty alongside their distinctive work approaches. The research demonstrates an important educational requirement to maintain AI benefits while preserving independent writing practices.

4.5. Interpretation and Implications

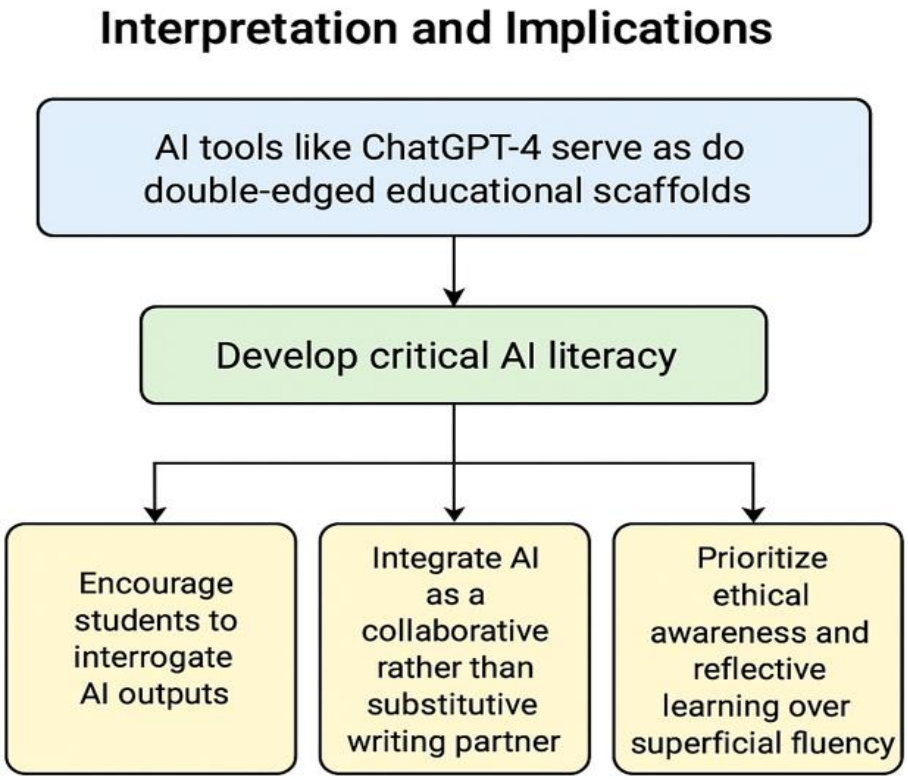

Figure 5.
Interpretation and pedagogical implications of using AI tools in academic writing.

Figure 5 depicts a conceptual flowchart which interprets and shows educational implications regarding the utilization of AI tools such as ChatGPT-4 for academic composition. The initial layer of the diagram introduces AI tools as “double-edged educational scaffolds” to show both positive and negative effects on student learning. The key recommendation calls for establishing critical AI literacy as the fundamental approach to address the dual aspects of this situation. The central principle produces three educational strategies that extend from it. Educational practice should lead students towards a habit of inquiry by challenging AI-produced outputs instead of having them accept such content without question. AI technology needs to function as an assistant tool instead of competing with human authors because its purpose should be to support human writers in their creative processes and logical thinking abilities. Both students and educators need to value ethical perspectives together with reflective learning above superficial fluency to prevent students from exchanging depth and original writing quality for speed and polished completion.

5. Discussion

This study investigates how university students in Pakistan interact with advanced writing tools, focusing particularly on their experiences with ChatGPT-4. The results show that students often benefit from clearer organization and a more structured presentation of their arguments when using the tool. However, this gain in fluency sometimes comes at the expense of deeper critical engagement. In cases where students rely too heavily on AI-generated input, they struggle to articulate their own viewpoints and maintain originality in their work. This tension between improved structure and reduced analytical depth highlights a broader concern about the role of such technologies in academic contexts. The findings align with recent discussions in higher education regarding the integration of AI into student learning and writing practices [29, 30].

5.1. AI and Structural Proficiency: Organizational Support or Mechanical Output?

The swift advantage of AI writing support consisted of enhanced structural arrangement in written materials. Individuals who employed ChatGPT-4 achieved superior results on thesis clarity assessment (M = 4.5), followed by paragraph coherence evaluation (M = 4.3). The tool thus proved helpful for academic writing development. Song and Song [31] showed that AI-enhanced writing software produced better-structured essays with improved coherence, specifically among ESL users. Structural organization challenges remain long-term obstacles for novice academic writers in non-native English contexts [26].

The essays using AI assistance during this research displayed typical patterns of argumentative writing. These assignments incorporated thesis declarations combined with orderly paragraph organization, topic statements, and seamless transitional elements. ChatGPT-4 functions as a structural coach that demonstrates high consistency in generating surface elements of academic discourse. The reception of academic materials depends heavily on "surface-level fluency," which Feng and Guo [18] define as crucial for evaluation purposes. This application makes AI serve as a digital workshop [21] that demonstrates basic academic communication structures through straightforward examples.

Students should use these structural enhancements with careful consideration. While organizational improvements enhance readability and clarity, they might not demonstrate profound intellectual engagement or authentic creativity [3]. The focus on formatting in AI-generated compositions often distanced writing from substantive meaning, producing sophisticated text without deep conceptual depth. ChatGPT-4's ability to quickly create academic text produces a deceptive effect that confuses surface mastery for weak content comprehension.

The writing tool sets an important foundation for academic writing mechanics but should not substitute for independent thought and analytical abilities. Experts agree that writing structure remains essential for author success, but the process must connect to advanced thinking abilities involving evidence evaluation, complex subject presentation, and argument response methods for genuine academic development [5]. The findings about critical engagement expose the evident conflict between form and substance, which will be analyzed in the upcoming section.

5.2. Diminished Critical Engagement and Intellectual Autonomy

The writing assistance tool showed improved results with formal elements, yet demonstrated reduced ability to promote critical thinking. Students using ChatGPT-4 achieved average scores of 2.9 on critical engagement tests, while students writing independently averaged 3.6. Literature examines generative AI criticisms because this technology produces academic language with ease, yet fails to create writing demonstrating sophisticated reasoning or contextual argumentation skills [32]. AI-generated texts might demonstrate professional writing style with strong coherence, yet they lack the academic depth expected in scholarly texts.

Our study demonstrated that AI-generated essays omitted essential critical thinking components upon close examination. The writing neglected to challenge fundamental premises, disregarded counterarguments, and failed to develop self-generated thought sequences. This pattern corresponds with Jakesch et al. [33] description of AI-generated text as "plausible-sounding but shallow" content that creates false perceptions of sophistication. Several participants adopted an uncritical approach to ChatGPT-4 outputs, integrating responses as complete essays without personal refinement. Students lose knowledge-construction agency when excessively depending on ChatGPT-4, diminishing their epistemic autonomy.

Students' inability to confront understanding problems through deliberate cognitive efforts raises significant concerns. Advanced academic learning requires metacognitive engagement – actively grappling with complex ideas while refining views through diverse perspectives. AI-based composition systems enable students to bypass essential cognitive processes when used for extensive support. AI's capacity for instant responses leads students to abandon critical reflection processes fundamental to deep learning progress [34].

A significant educational risk emerges from this development. According to Aoun [5] education must maintain emphasis on abilities AI cannot replicate, including critical thinking, creative thinking, and ethical reasoning. Student use of writing software to generate content risks eroding essential cognitive development. Developing students' skills to approach AI tools requires explicit instruction about reflective use rather than cognitive replacement. Students need guidance to prevent AI tools from fostering passive learning cultures while preserving intellectual independence and educational growth.

5.3. Source Integration: A False Sense of Academic Rigor?

One major discovery of this research analyzed how students incorporated sources into their papers. AI-assisted writing resulted in peer-reviewed citations comprising 75% of bibliographies, compared to 32% in non-AI work. While tools like ChatGPT-4 help students obtain scholarly materials and emulate academic writing standards, a critical deficiency emerged: only 37% of AI-assisted students verified sources, versus 85% of non-AI students.

This gap reveals a troubling pattern in engagement with AI-generated citations. Students displayed unwarranted confidence in ChatGPT-4's references, neglecting authenticity checks that could compromise academic standards. These results align with studies warning about LLMs generating "hallucinated" citations that appear credible but reference nonexistent publications [35]. Unverified hallucinations diminish academic validity.

The educational gap stems from instructors presenting citations as mechanical rather than analytical. Academic excellence requires source interaction, legitimacy evaluation, and argument development—not mere insertion. Students risk mistaking formally styled references for scholarly rigor.

Overreliance on machine-generated citations threatens essential research capabilities: source assessment and synthesis. While AI can identify relevant research, students must critically analyze outputs through independent verification. Teaching critical AI literacy should emphasize proactive information verification and responsible scholarly integration.

5.4. Student Perceptions: Efficiency, Dependency, and the Ambiguity of Assistance

Survey results revealed divergent perceptions of AI tools. AI-assisted students rated efficiency highly (M = 4.5), valuing rapid drafting and usability in time-sensitive contexts. However, they simultaneously expressed dependency concerns (M = 3.8 on "I worried about dependency on AI"), reflecting ambivalence about AI's impact on academic development [1].

Non-AI participants reported higher perceived autonomy (M = 4.2) and deeper content understanding. AI's efficiency gains may foster artificial confidence that detaches students from essential cognitive processes [36]. While facilitating text production, these tools often fail to cultivate sustained engagement or reflective thinking.

Educational institutions must establish pedagogical principles protecting intellectual growth when integrating AI [37]. Tools should assist—not replace—the mental work of composition. Without clear frameworks, students risk becoming passive content consumers rather than active knowledge builders.

5.5. Ethical Considerations and Academic Integrity

Higher education faces ethical quandaries regarding authorship rights and integrity norms. Participants expressed mixed feelings: some appreciated ChatGPT-4's clear outputs, while others noted confusion about boundaries between assistance and misconduct. This discomfort mirrors academic debates linking AI use to compromised student contribution [38].

The core challenge lies in distinguishing legitimate assistance from inappropriate substitution. Students experience ethical confusion when delegating ideation and exposition to systems producing machine-generated content. Jakesch et al. [33] argue this shift undermines education's core mission: developing analytical and reflective abilities through writing.

Both students and educators often mistakenly assume that AI-generated content is inherently reliable. Our study found instances of unverified incorporation of ChatGPT content. While the AI group cited more peer-reviewed sources, many were inaccurate or fabricated due to "AI hallucination"[1], undermining academic standards.

Institutions must develop explicit ethical AI frameworks rather than relying solely on punitive measures. Students require an understanding of academic values beyond rule compliance, McCabe [39]. Garcia Ramos and Wilson-Kennedy [20] advocate shifting from enforcement to cultivating critical digital ethics awareness.

Addressing accessibility disparities is equally crucial. Students with superior digital access and English proficiency gain disproportionate AI advantages, potentially exacerbating educational inequities. Data literacy education should cover AI's development process, factual verification, and algorithmic fairness assessment across diverse user perspectives.

5.6. Pedagogical Implications

This study offers valuable insights for teacher education, particularly in the context of academic writing instruction in AI-driven learning environments (see Figure 5). The findings suggest that while tools like ChatGPT-4 are effective in helping students organize their thoughts and improve text coherence, they fall short in fostering higher-order cognitive engagement. The essays produced by students with AI support demonstrated clear structural organization, but their critical thinking skills and intellectual autonomy were noticeably weaker. Previous research supports these findings, indicating that while generative AI excels at replicating academic writing conventions, it struggles to cultivate deep, context-driven [40].

Educators face a significant challenge in finding ways to integrate technology into the classroom without diminishing students' critical thinking skills and academic independence. A practical solution is to design scaffolded writing tasks where technology serves as a tool for initiating ideas rather than replacing the students' intellectual engagement. Teaching staff can request that students use ChatGPT-4 to produce initial drafts and outlines, while students should then conduct activities that require them to assess and modify the AI-generated content. Teaching approaches such as finding rhetorical gaps, along with

counterargument presentation and personal insight addition in writing, can allow students to benefit from AI guidance structures while practicing metacognitive reflection.

The analysis demonstrates the immediate requirement to teach students better skills regarding information literacy when AI aids their writing tasks. Several students confirmed accepting AI-generated citations without validating their authenticity leading to substantial issues regarding academic research credibility, according to data. The results demonstrate AI's known capability to produce fake citations, whereas Marcus and Davis [35] have recently warned about this issue. Writing instruction must provide students with specific training that teaches how to authenticate AI-recommended references and how to verify sources through academic databases and enables learners to comprehend the epistemological problems of using generative content [41].

Academic curricula need to incorporate critical AI literacy components for student education. AI systems learn from previous data, which leads them to perpetuate mainstream understandings while excluding different or culturally unique perspectives [38]. The educational process should challenge students to reveal ideological assumptions within AI-generated text while verifying whether its structures adhere to various epistemic traditions.

Finally, teacher preparedness remains crucial. Educators often lack confidence in addressing AI issues due to insufficient training and institutional support, Atabek [6]. Atabek [6] emphasizes that ongoing professional development and departmental collaboration with administrative endorsement are essential to bridge the gap between technological innovation and classroom implementation.

6. Conclusion

This study examined how the use of technological tools influences undergraduate students' academic essay writing. The aim was not only to assess whether students could write more fluently with technological assistance but also to understand how these tools impact deeper learning, critical thinking, and independent reasoning. Using a randomized controlled trial and analyzing student essays through Toulmin's model, the findings offer a detailed and nuanced understanding of the effects on students' writing and cognitive processes.

Students using AI-generated work produced pieces with better overall writing organization. Their academic work showed organization since each essay contained well-defined thesis statements together with coherent paragraph transitions. The solid structural quality of tools such as ChatGPT-4 might provide assistance to students who need help with thought organization or writing in formal academic styles. The results changed when we examined the underlying characteristics. The student essays written with AI assistance usually contained superficial content. Students produced essays with standardized arguments and predictable reasoning patterns, which showed minimal signs of facing complex ideas or counterarguments. Although the content structure held strong, the thinking aspects felt hollow.

The written work produced by students who did not use AI revealed greater scholarship with reduced professional quality. The arguments demonstrated proof of mental effort but simultaneously showed signs of intensive thought work. They examined the trustworthiness of their sources while considering real-world relevance and providing their own personal insights for evaluation purposes. Their writing contained more errors; however, it displayed enhanced intellectual independence with a real comprehension of the material.

In general, perceptions of students maintained recurring patterns about their writing process. Users who received assistance from AI systems found the tools simple to use with fast processing times. The tool enabled writers to compose their papers quickly while gaining stronger control over their document outlines. The tool received mixed reactions from students who debated both their probable dependence on artificial intelligence and doubts about their understanding of the subject matter. Students admitted feelings of guilt because they doubted whether the AI usage violated ethics. The students who wrote on their own experienced tough writing situations that brought them satisfaction. Students who did the work by hand reported that confronting assignments independently improved their ability to think critically and compose better research questions to achieve a clearer understanding of their subjects.

Education faces its most critical challenge in this present day. The benefits of Artificial Intelligence tools become evident because they assist students with various academic tasks while reducing obstacles, which enables students to plan, draft, and revise their work. The systems pose significant hazards whenever students employ them without proper guidance. Students who depend on AI for their assignments will likely skip essential learning steps during the process. The creation of clean paragraphs by itself does not represent the goal of writing because the main objectives involve idea creation, alongside evidence analysis, as well as mastering complex subject matter.

The resolution between AI tool restrictions and unbridled classroom access should not be absolute bans or complete freedoms. The need exists to establish a comprehensive method for AI utilization in education that promotes both competent use and original comprehension, as well as the maintenance of independent work. Students can learn about AI tool utility from their teachers by using these tools as introduction points rather than concluding points. Students should take advantage of AI tools for brainstorming and draft refinement, yet receive lessons about verifying generated content and the development of independent thinking.

The study demonstrates the need for teachers and students to establish a harmonious approach between AI assistance and independent critical thought. The process of genuine learning occurs through intellectual struggles to tackle concepts, test beliefs, and execute work autonomously. Teaching students to perceive AI as a tool that supports their cognitive development will allow us to maximize the advantages of both efficient processes and authentic authorship that remain creatively unadulterated.

6.1. Limitations and Directions for Future Research

The study offers valuable insights into AI-assisted writing practices, but its findings are limited due to specific constraints. The sample size, which consists of 50 undergraduate students from a single institution in Pakistan, restricts the generalizability of the results. Future research should involve multiple institutions across diverse linguistic and cultural contexts to strengthen the findings, as the randomized controlled design significantly enhances validity. It is also important to give special attention to student experiences with technology, as their use of such tools evolves based on their educational training and available technological resources [30].

This research examines mostly short-duration writing responsibilities. Following continuous student exposure to AI tools throughout development, researchers must conduct repeated study measurements. Future research should explore how students evolve with AI assistance by studying changes in dependence on AI tools throughout their writing development. The research should also study whether developing a habit of depending on AI tools prevents writing skill growth. Extended research-based studies need to evaluate both the curriculum-related threats as well as the advantages of integrating AI technology [42].

Research must investigate in detail how students react emotionally when they utilize AI writing assistance. Several participants in the study perceived AI as a beneficial partner, yet others experienced negative emotions due to AI assistance. Investigating how students handle their emotions towards AI technology provides a better understanding of the psychological effects of artificial intelligence implementations in educational settings. Researchers should perform qualitative inquiries by conducting interviews and focus groups to track students' progressive interactions with AI tools.

References

[1] H. Alkaissi and S. I. McFarlane, "Artificial hallucinations in ChatGPT: Implications in scientific writing," *Cureus*, vol. 15, no. 2, 2023. https://doi.org/10.7759/cureus.35179

[2] M. Al-Raimi, B. A. Mudhsh, M. H. Muqaibal, and Y. Al-Yafaei, "To what extent has artificial intelligence impacted EFL teaching and learning? A systematic review," *JURNAL ARBITRER*, vol. 11, no. 3, pp. 399-412, 2024. https://doi.org/10.25077/ar.11.3.399-412.2024

[3] M. Al-Raimi, B. A. Mudhsh, Y. Al-Yafaei, and S. Al-Maashani, "Utilizing artificial intelligence tools for improving writing skills: Exploring Omani EFL learners' perspectives," in *Forum for Linguistic Studies (Transferred)*, 2024, vol. 6, no. 2, pp. 1177-1177.

[4] L. W. Anderson and D. R. Krathwohl, *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. New York, NY, USA: Longman, 2001.

[5] J. E. Aoun, *Robot-proof: Higher education in the age of artificial intelligence*. Cambridge, MA, USA: MIT Press, 2017.

[6] O. Atabek, "Challenges in integrating technology into education," *arXiv preprint arXiv:1904.06518*, 2019.

[7] A. Barrett and A. Pack, "Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 59, 2023. https://doi.org/10.1186/s41239-023-00427-0

[8] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big??," in *In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623)*, 2021.

[9] O. Boateng and B. Boateng, "Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education," *World Journal of Advanced Research and Reviews*, vol. 25, no. 1, pp. 2012–2017, 2025. https://doi.org/10.30574/wjarr.2025.25.1.0253

[10] A. Borji and R. Mohammadian, "Artificial creativity: A comprehensive review," *Artificial Intelligence Review*, vol. 56, no. 2, pp. 123–167, 2023. https://doi.org/10.1007/s10462-023-10478-9

[11] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77-101, 2006. https://doi.org/10.1191/1478088706qp063oa

[12] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," presented at the In Conference on fairness, accountability and transparency (pp. 77-91). PMLR, 2018.

[13] A. Burgess-Proctor, "Methodological and ethical issues in feminist research with abused women: Reflections on participants' vulnerability and empowerment," in *Women's Studies International Forum*, 2015, vol. 48, pp. 124-134.

[14] C. K. Y. Chan, "A comprehensive AI policy education framework for university teaching and learning," *International journal of educational technology in higher education*, vol. 20, no. 1, p. 38, 2023.

[15] J. Cohen, "A cofficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37-46, 1960. https://doi.org/10.1177/001316446002000104

[16] J. W. Creswell, *Designing and conducting mixed methods research*. Thousand Oaks, CA, USA: SAGE Publications, 2018.

[17] H. Dahlström, "Digital writing tools from the student perspective: Access, affordances, and agency," *Education and Information Technologies*, vol. 24, no. 2, pp. 1563-1581, 2019. https://doi.org/10.1007/s10639-018-9844-x

[18] R. Feng and Q. Guo, "Second language speech fluency: what is in the picture and what is missing," *Frontiers in Psychology*, vol. 13, p. 859213, 2022. https://doi.org/10.3389/fpsyg.2022.859213

[19] A. Field, *Discovering statistics using IBM SPSS statistics*. London, UK: SAGE Publications, 2018.

[20] J. Garcia Ramos and Z. Wilson-Kennedy, "Promoting equity and addressing concerns in teaching and learning with artificial intelligence," *Frontiers in Education*, vol. 9, p. 1487882, 2024. https://doi.org/10.3389/feduc.2024.1487882

[21] J. P. Gee, *Situated language and learning: A critique of traditional schooling*. New York, NY, USA: Routledge, 2004.

[22] M. Gerlich, "AI tools in society: Impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, p. 6, 2025. https://doi.org/10.3390/soc15010006

[23] D. Muijs, *Doing quantitative research in education with SPSS*. London, UK: Sage, 2010.

[24] S. E. Toulmin, *The uses of argument*. Cambridge, UK: Cambridge University Press, 2003.

[25] J. OpenAI Achiam *et al.*, "GPT-4 technical report. arXiv," *arXiv preprint arXiv:2303.08774*, 2023.

[26] K. Hyland, *Second language writing*. Cambridge university press, 2019.

[27] S. Wilson, *Thinking differently about leadership: A critical history of leadership studies*. Cheltenham, UK: Edward Elgar Publishing, 2016.
[28] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learning and instruction*, vol. 4, no. 4, pp. 295-312, 1994.
[29] M. Khalifa and M. Albadawy, "Using artificial intelligence in academic writing and research: An essential productivity tool," *Computer Methods and Programs in Biomedicine Update*, vol. 5, p. 100145, 2024. https://doi.org/10.1016/j.cmpbup.2024.100145
[30] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education–where are the educators?," *International journal of educational technology in higher education*, vol. 16, no. 1, pp. 1-27, 2019. https://doi.org/10.1186/s41239-019-0171-0
[31] C. Song and Y. Song, "Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students," *Frontiers in Psychology*, vol. 14, p. 1260843, 2023. https://doi.org/10.3389/fpsyg.2023.1260843
[32] A. Nguyen, Y. Hong, B. Dang, and X. Huang, "Human-AI collaboration patterns in AI-assisted academic writing," *Studies in Higher Education*, vol. 49, no. 5, pp. 847-864, 2024. https://doi.org/10.1080/03075079.2024.2323593
[33] M. Jakesch, J. T. Hancock, and M. Naaman, "Human heuristics for AI-generated language are flawed," *Proceedings of the National Academy of Sciences*, vol. 120, no. 11, p. e2208839120, 2023.
[34] M. Shahbaz, M. Rauf, and S. Rehman, "Effects of AI-based writing tools on Pakistani university students' academic writing skills," *International Journal of Educational Research*, vol. 124, p. 102081, 2025. https://doi.org/10.1016/j.ijer.2024.102081
[35] G. F. Marcus and E. Davis, *Rebooting AI: Building artificial intelligence we can trust*. New York, NY, USA: Pantheon Books, 2019.
[36] A. R. Malik, Y. Pratiwi, K. Andajani, I. W. Numertayasa, S. Suharti, and A. Darwis, "Exploring artificial intelligence in academic essay: Higher education student's perspective," *International Journal of Educational Research Open*, vol. 5, p. 100296, 2023. https://doi.org/10.1016/j.ijedro.2023.100296
[37] N. Lacković, "Postdigital living and algorithms of desire," *Postdigital Science and Education*, vol. 3, pp. 280-282, 2021. https://doi.org/10.1007/s42438-020-00141-4
[38] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI chatbots in education: systematic literature review," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 56, 2023. https://doi.org/10.1186/s41239-023-00426-1
[39] D. McCabe, "Cheating and honor: Lessons from a long-term research project," *Second handbook of academic integrity*, pp. 599-610, 2024. https://doi.org/10.1007/978-981-287-079-7_35-1
[40] R. Pereira, I. W. Reis, V. Ulbricht, and N. d. Santos, "Generative artificial intelligence and academic writing: An analysis of the perceptions of researchers in training," *Management Research: Journal of the Iberoamerican Academy of Management*, vol. 22, no. 4, pp. 429-450, 2024. https://doi.org/10.1108/MRJIAM-01-2024-1501
[41] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," presented at the In Proceedings of the conference on Fairness, Accountability, and Transparency (pp. 59-68), 2019.
[42] S. A. Popenici and S. Kerr, "Exploring the impact of artificial intelligence on teaching and learning in higher education," *Research and practice in technology enhanced learning*, vol. 12, no. 1, p. 22, 2017. https://doi.org/10.1186/s41039-017-0062-8