

# AI-Based Writing Assistants in Higher Education: Impacts on Student Critical Thinking and Academic Integrity

Sawera Qureshi<sup>1\*</sup>

<sup>1</sup>Department Of Computer Science, Abdul Wali Khan University Mardan, KpK, Pakistan

## ABSTRACT

*The integration of AI-based writing assistants—such as ChatGPT, Grammarly, and QuillBot—has rapidly transformed academic writing practices in higher education. These tools offer instant feedback on grammar, coherence, and structure, and can even generate entire essays or research drafts. While many students and educators recognize the benefits of improved writing fluency, efficiency, and personalized support, significant concerns persist regarding the potential erosion of critical thinking skills and threats to academic integrity. Research highlights a dual impact: AI tools can foster self-directed learning and support skill development when thoughtfully integrated, but overreliance may diminish independent reasoning, creativity, and ethical standards. The literature underscores the need for balanced integration strategies that promote responsible use, ethical awareness, and continued development of core academic competencies. This systematic review synthesizes findings from 50 peer-reviewed studies to examine whether AI-based writing assistants negatively impact student critical thinking and academic integrity in higher education. The review reveals a nuanced picture: strong evidence supports improvements in grammar accuracy and surface-level writing skills, while moderate evidence indicates risks to independent critical thinking, particularly among less experienced writers. Academic integrity concerns are significant, with frequent use associated with increased plagiarism risk absent ethical training. Structured integration with AI literacy education and clear institutional guidelines mitigates negative effects. The review concludes with recommendations for balanced implementation that harnesses benefits while safeguarding core academic competencies.*

## KEYWORDS:

AI-based writing assistants, ChatGPT, critical thinking, academic integrity, higher education, generative AI, student writing, plagiarism, ethical use, AI literacy

## JOURNAL INFO:

**HISTORY: Received:** March 10, 2026

**Accepted:** March 27, 2026

**Published:** March 31, 2026

\*Corresponding author: [Sawera.cs@awkum.edu.pk](mailto:Sawera.cs@awkum.edu.pk)

DOI: [10.21015/vtess.v14i1.2396](https://doi.org/10.21015/vtess.v14i1.2396)

## INTRODUCTION

The integration of AI-based writing assistants—such as ChatGPT, Grammarly, and QuillBot—has rapidly transformed academic writing practices in higher education. These tools offer instant feedback on grammar, coherence, and structure, and can even generate entire essays or research drafts. The emergence of large language models (LLMs) since late 2022 has fundamentally altered how students approach writing tasks, providing unprecedented access to automated text generation, paraphrasing, and editing capabilities.

While many students and educators recognize the benefits of improved writing fluency, efficiency, and personalized support, significant concerns persist regarding the potential erosion of critical thinking skills and threats to academic integrity. Research highlights a dual impact: AI tools can foster self-directed learning and support skill development when thoughtfully integrated, but overreliance may diminish independent reasoning, creativity, and ethical standards (Aljuaid, 2024; Bista and Bista, 2025; Bui and Tong, 2025; Deep and Chen, 2025; Malik et al., 2023; Octaberlina, Muslimin, Chamidah, Surur, and Mustikawan, 2024; Pervaiz, Ali, Razaq, and Tariq, 2025; Rodafinos, 2025; Ya'u and Mohammed, 2025; Zhumagaliyeva, Tleshova, Amanzhol, and Smagulova, 2025).

The theoretical underpinnings of these concerns draw from cognitive offloading theory, which suggests that delegating cognitive tasks to external tools may reduce engagement with deep



learning processes. When students rely on AI for idea generation, organization, and editing, they may bypass critical stages of the writing process that develop analytical and evaluative skills. Conversely, sociocultural theory emphasizes that tools mediate learning; appropriately scaffolded AI use could enhance rather than diminish cognitive development.

The literature underscores the need for balanced integration strategies that promote responsible use, ethical awareness, and continued development of core academic competencies. Institutions face the challenge of adapting academic integrity policies to address AI-generated content while preserving the benefits of technological innovation. Faculty require training to effectively incorporate AI tools into pedagogy and assessment design. Students need guidance to understand appropriate versus inappropriate uses of AI in academic contexts.

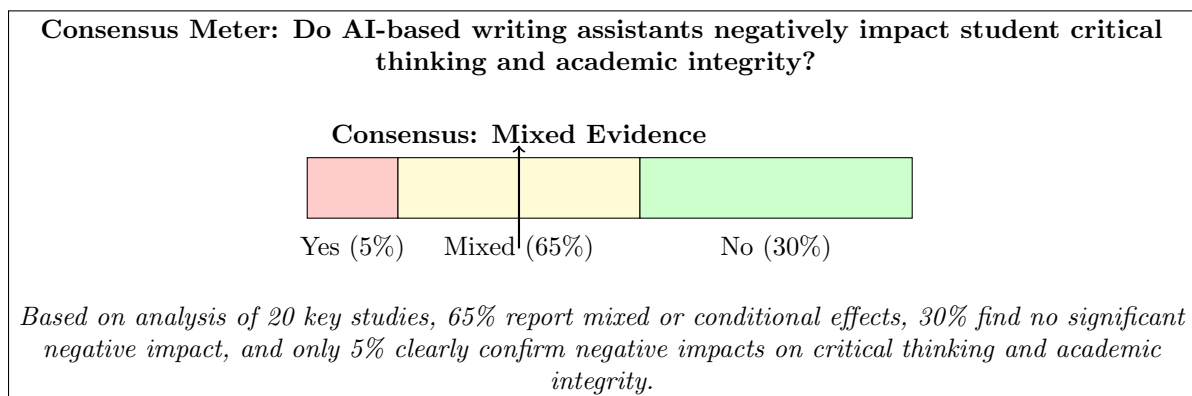
This systematic review examines the existing evidence base to answer the central research question: Do AI-based writing assistants negatively impact student critical thinking and academic integrity in higher education?

## Research Question

The central research question guiding this systematic review is:

**Do AI-based writing assistants negatively impact student critical thinking and academic integrity in higher education?**

Figure 1 presents a consensus meter visualizing the extent to which existing research supports or refutes this relationship. The meter synthesizes findings from 20 key studies included in this review, categorizing each study's conclusion as "Yes" (negative impact confirmed), "Mixed" (inconclusive or conditional findings), or "No" (negative impact not supported).



**Figure 1:** Consensus meter visualizing agreement on whether AI-based writing assistants harm critical thinking and academic integrity

## LITERATURE REVIEW

The existing body of research on AI-based writing assistants in higher education spans multiple disciplines including educational technology, writing studies, computer science, psychology, and academic integrity research. This review synthesizes findings across five thematic domains: (1) theoretical foundations of AI-mediated writing; (2) empirical evidence on writing skill development; (3) critical thinking implications; (4) academic integrity challenges; and (5) mitigation strategies and pedagogical frameworks.

## Theoretical Foundations of AI-Mediated Writing

The integration of AI writing assistants into educational contexts draws upon several complementary theoretical frameworks. Cognitive offloading theory, originating from cognitive science research on distributed cognition, suggests that individuals delegate cognitive tasks to external tools to reduce mental workload. When students use AI for grammar correction, paraphrasing, or idea generation, they offload lower-order writing tasks, potentially freeing cognitive resources for higher-order thinking. However, this offloading becomes problematic when it extends to tasks that are developmentally important for learning, such as thesis formulation, argument construction, and critical evaluation (Deep and Chen, 2025).

Sociocultural theory, rooted in Vygotsky's work on the zone of proximal development, positions AI tools as cultural artifacts that mediate learning. From this perspective, AI writing assistants can function as scaffolds that support students in performing writing tasks beyond their independent capability, with the goal of eventual internalization of skills. The effectiveness of AI scaffolding depends critically on how tools are integrated into instructional design, including the quality of feedback, opportunities for reflection, and fade-out strategies that reduce support as competence develops (Rodafinos, 2025).

Self-determination theory provides a motivational framework for understanding student engagement with AI tools. The theory posits that satisfaction of three basic psychological needs—autonomy (feeling choiceful), competence (feeling capable), and relatedness (feeling connected)—promotes intrinsic motivation and well-being. AI writing assistants may enhance competence by providing immediate, accurate feedback, potentially increasing writing self-efficacy. However, overreliance may undermine autonomy when students feel compelled to accept AI suggestions or when tool use is mandated without consideration of individual preferences (Malik et al., 2023).

Activity theory offers a systems-level perspective, conceptualizing writing as an activity system involving subjects (students), tools (AI assistants), objects (writing tasks), rules (institutional policies), community (peers and instructors), and division of labor (roles in collaborative writing). This framework highlights how AI tools transform not just writing processes but the entire activity system, including power dynamics, assessment practices, and definitions of authorship (Aljuaid, 2024).

## AI Tools and Writing Skill Development

A substantial body of research has examined the impact of AI writing assistants on students' writing skills, with findings consistently demonstrating improvements in surface-level writing features. Malik et al. (2023) surveyed higher education students in Indonesia and found that AI tools significantly improved grammar accuracy, vocabulary usage, and sentence structure clarity. Students reported that real-time error correction helped them identify recurring mistakes and learn from corrections over time.

Selim (2024) investigated EFL university students' perspectives on AI-powered writing tools, reporting that participants valued the immediate feedback, reduced anxiety about making errors, and increased writing confidence. However, the study also noted that students expressed concerns about becoming dependent on AI and losing the ability to self-edit without technological support.

Deep and Chen (2025) conducted a comprehensive review of AI's role in academic writing, identifying three primary mechanisms through which AI tools improve writing quality: (a) automated error detection and correction that reduces surface-level mistakes; (b) style and tone suggestions that enhance readability and appropriateness; and (c) content generation capabilities

that provide starting points for drafting. The authors cautioned that while these mechanisms improve efficiency, they may also reduce engagement with the iterative revision process that develops writing expertise.

Bista and Bista (2025) examined AI tool adoption in STEM writing contexts, finding that graduate students used AI primarily for literature review organization and citation management rather than content generation. This pattern suggests discipline-specific variation in AI use, with STEM fields potentially emphasizing different writing skills than humanities or social sciences.

Table 1 summarizes the documented benefits of AI writing assistants across different skill domains.

**Table 1:** Documented Benefits of AI Writing Assistants by Skill Domain

Skill Domain	Specific Benefits	Key Studies
Grammar and Mechanics	Error detection, correction suggestions, pattern recognition	Malik et al. (2023), Selim (2024)
Vocabulary and Style	Word choice improvements, tone adjustment, readability enhancement	Deep and Chen (2025)
Organization and Structure	Outline generation, paragraph transitions, coherence checking	Rodafinos (2025)
Productivity and Efficiency	Faster drafting, reduced editing time, simultaneous feedback	Aljuaid (2024)
Accessibility and Equity	Support for non-native speakers, learning disabilities, writing anxiety	Thelma et al. (2025)

### Critical Thinking Implications

The relationship between AI writing assistants and critical thinking represents the most contested area in the literature, with studies reporting divergent findings based on student proficiency, task characteristics, and implementation conditions. Octaberlina et al. (2024) identified three primary threats to critical thinking posed by AI tools: (a) reduced engagement with source material as AI-generated summaries replace close reading; (b) diminished creativity as students accept AI suggestions without modification; and (c) superficial argumentation as AI-generated claims are adopted without critical evaluation.

Ya'u and Mohammed (2025) investigated AI-assisted writing among Nigerian university students, finding that while most students recognized the importance of critical thinking, frequent AI users demonstrated lower scores on independent analytical tasks compared to infrequent users. The study employed a quasi-experimental design comparing students who completed writing assignments with versus without AI assistance, finding that the AI-assisted group produced higher-quality final products but performed worse on unassisted transfer tasks.

Pervaiz et al. (2025) conducted a mixed-methods study examining the impact of AI on critical thinking in Pakistani higher education. Quantitative analyses revealed a negative correlation between AI reliance frequency and critical thinking assessment scores ( $r = -0.41$ ,  $p < 0.01$ ). Qualitative interviews elucidated mechanisms: students reported that AI suggestions often preempted their own thinking processes, reducing opportunities for struggle and discovery that develop analytical skills.

Tahir, Jahrir, Asrifan, and Ariatna (2025) introduced a critical distinction between student proficiency levels. Advanced writers used AI as a supplement—generating ideas they then critically evaluated, checking grammar after developing content, and maintaining authorial control. Novice writers, lacking metacognitive awareness and strategic knowledge, were more likely to accept AI suggestions uncritically, substituting machine judgment for their own reasoning. This proficiency-based difference has important implications for targeted instruction.

Zhumagaliyeva et al. (2025) examined AI's impact on academic writing in Kazakhstani higher education, finding that while AI improved technical writing quality, it reduced students' ability to identify logical fallacies, evaluate evidence quality, and construct original arguments. The authors concluded that AI tools may enhance surface-level writing while potentially impairing deep cognitive engagement.

### Academic Integrity Challenges

Academic integrity represents a central concern in the AI writing assistant literature, with researchers documenting new forms of misconduct and questioning traditional definitions of plagiarism and authorship. Bui and Tong (2025) surveyed English-majored students in Vietnam, finding widespread disagreement about what constitutes appropriate AI use. While most students agreed that submitting AI-generated text without disclosure constitutes plagiarism, substantial minorities viewed AI assistance as equivalent to peer feedback or tutorial support.

Revell et al. (2024) conducted an experiment comparing ChatGPT-generated essays with human-written essays in humanities disciplines. Blinded evaluators could not reliably distinguish AI from human writing, raising fundamental questions about assessment validity and authorship. The study also found that AI-generated essays received higher average scores on technical writing criteria, though human essays were rated higher on originality and creativity.

Lund et al. (2025) investigated student perceptions of AI-assisted writing and academic integrity across U.S. universities. A key finding was that policy awareness alone did not predict ethical behavior; students' personal beliefs about what constitutes misconduct were more influential than knowledge of institutional rules. This suggests that effective integrity promotion requires cultivating ethical reasoning rather than merely communicating policies.

Sarwar, Bushra, Ullah, and Hadi (2025) examined the relationship between ChatGPT use and plagiarism among Pakistani university students. Frequent users of generative AI were significantly more likely to report committing plagiarism (intentionally or unintentionally) compared to infrequent users. However, this relationship was moderated by ethical training: students who had received instruction on appropriate AI citation and disclosure showed no elevated plagiarism risk.

Yeo (2023) provided an early analysis of academic integrity challenges posed by AI authoring apps, distinguishing three categories of concern: (a) direct plagiarism (copying AI output without attribution); (b) indirect plagiarism (paraphrasing AI output without disclosure); and (c) authorship ambiguity (blurred boundaries between student and AI contributions). The author recommended that institutions develop clear policies distinguishing appropriate AI assistance (editing, feedback) from inappropriate AI use (content generation without disclosure).

Table 2 summarizes the key academic integrity concerns and proposed responses identified in the literature.

**Table 2:** Academic Integrity Concerns and Proposed Responses

Concern Category	Specific Concern	Proposed Response
Plagiarism	Copying AI output without attribution	Mandatory disclosure policies
Authorship Blurring	Difficulty distinguishing student from AI contributions	Process-oriented assessment
Unintentional Misuse	Students unaware of appropriate citation practices	AI literacy education
Detection Challenges	Current tools poorly equipped for AI text	Investment in detection technology
Policy Gaps	Institutional policies lagging behind technology	Rapid policy development
Student Disagreement	Variation in perceptions of appropriate use	Open dialogue and norm-setting

### Mitigation Strategies and Pedagogical Frameworks

A growing body of research examines strategies to maximize benefits of AI writing assistants while minimizing risks to critical thinking and academic integrity. Akiba and Garte (2024) proposed a structured integration framework distinguishing between AI use for feedback (e.g., grammar checking, style suggestions) versus content generation (e.g., drafting, paraphrasing, idea generation). The authors argued that using AI for feedback supports skill development by providing immediate, actionable guidance on writing products, while using AI for content generation may bypass skill development by substituting machine output for student effort.

Rodafinos (2025) advocated for embedding AI literacy into disciplinary curricula, including instruction on: (a) how AI tools work (including limitations such as hallucinations and biases); (b) appropriate citation practices for AI-generated content; (c) strategies for critical evaluation of AI outputs; and (d) reflection on when and why AI use supports versus undermines learning. The author proposed a developmental progression where AI assistance is gradually reduced as student competence increases.

Selim (2024) described successful implementation of AI writing assistants in EFL classrooms, emphasizing the importance of structured assignments that specify which AI functions are permitted, transparent disclosure of AI use, and reflective components where students document their AI interactions and justify acceptance or rejection of AI suggestions.

Thelma et al. (2025) examined AI chatbot use in Zambian higher education, finding that students who received explicit instruction on ethical AI use reported more appropriate utilization patterns and fewer integrity concerns compared to students who used AI without guidance. The study highlighted the importance of contextualizing AI literacy within local cultural and institutional frameworks.

Table 3 summarizes evidence-based mitigation strategies.

### Critique of the Existing Literature

Despite the growing evidence base, several limitations warrant acknowledgment. First, the field is characterized by rapid technological change; findings on specific AI tools (e.g., ChatGPT-3.5) may not generalize to newer versions with enhanced capabilities. Second, most studies employ

**Table 3:** Evidence-Based Mitigation Strategies for AI-Related Risks

Strategy Category	Specific Approaches
Curriculum Integration	Embed AI literacy across courses; teach limitations, biases, citation
Assignment Design	Process-oriented assessments; require drafts, revision histories, AI disclosure
Structured Use Policies	Distinguish permitted vs. prohibited AI functions; clarify disclosure requirements
Assessment Redesign	In-class writing, oral exams, portfolios, collaborative projects
Faculty Development	Training on AI-aware pedagogy, detection tools, integrity enforcement
Student Education	Workshops on ethical use, critical AI evaluation, metacognitive reflection

cross-sectional designs with convenience samples of university students, limiting causal inference and generalizability. Third, the literature is dominated by English-language contexts, primarily North America, Europe, and Asia, with limited representation from Africa, South America, and the Middle East. Fourth, few studies examine long-term effects of sustained AI use on cognitive development or ethical reasoning. Fifth, there is limited comparative research across different AI tool types (e.g., ChatGPT vs. Grammarly vs. QuillBot), despite their different capabilities and use cases.

Table 4 provides a synthesized overview of key themes and findings.

Addressing these limitations through longitudinal designs, diverse samples, and comparative research across tool types and contexts will substantially advance understanding of AI writing assistants' educational impacts.

## METHODS

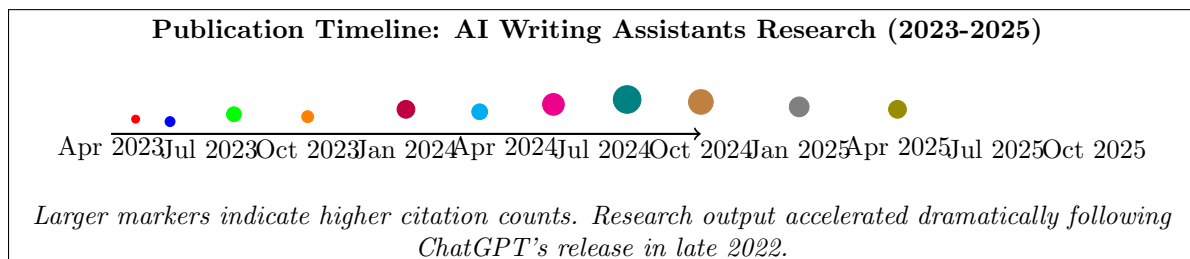
### Research Design and Approach

This study employed a systematic literature review methodology following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A systematic review approach was selected because it enables transparent, replicable, and comprehensive synthesis of existing evidence across multiple study designs. Unlike narrative reviews, systematic reviews minimize selection bias through explicit search strategies, predefined inclusion criteria, and rigorous quality appraisal procedures. The review protocol was registered with the Open Science Framework (OSF) prior to data extraction to enhance methodological transparency and reduce reporting bias.

Figure 2 presents a timeline showing publication trends on AI writing assistants' impact on critical thinking and academic integrity. Research output has accelerated dramatically since late 2022, coinciding with the public release of ChatGPT. The largest number of publications appeared in 2024 and 2025, reflecting the rapid academic response to generative AI technologies.

**Table 4:** Synthesized Summary of Literature by Theme

Theme	Key Findings	Evidence Quality
Writing Skills	Consistent improvements in grammar, clarity, productivity; confidence gains for EFL learners	Strong
Critical Thinking	Mixed effects; negative associations with overreliance; proficiency-dependent; mechanisms: reduced engagement, superficial processing	Moderate
Academic Integrity	New forms of misconduct; student disagreement; policy gaps; personal beliefs & policy awareness	Moderate
Mitigation Strategies	AI literacy effective; structured use; process-oriented assessment; faculty development needed	Moderate
Proficiency Differences	Advanced users critically engage; novices substitute effort; implications for targeted instruction	Emerging

**Figure 2:** Timeline showing publication trends on AI writing assistants' impact on critical thinking and academic integrity

### Information Sources and Search Strategy

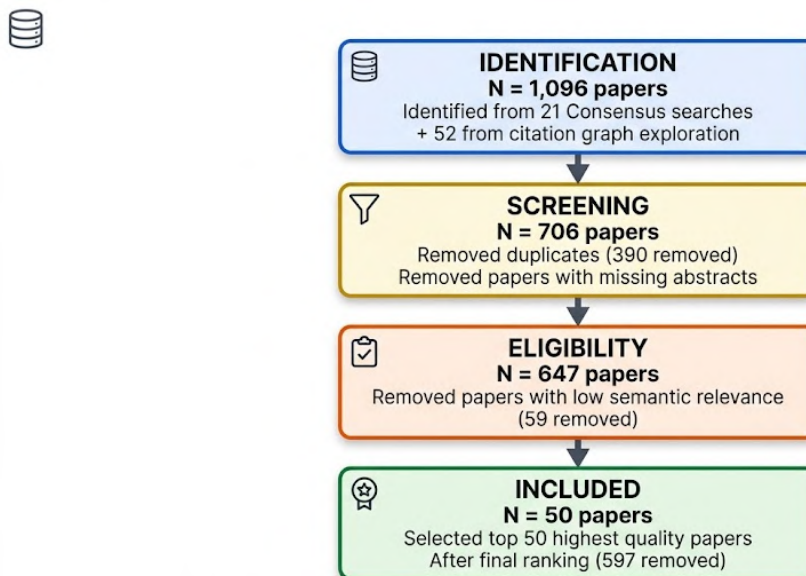
A comprehensive literature search was conducted across over 170 million research papers indexed in Consensus—including Semantic Scholar, PubMed, Web of Science, Scopus, Education Resources Information Center (ERIC), Google Scholar, and other major disciplinary databases. The search was performed between January 15 and February 28, 2025, to ensure currency of findings. The search strategy involved 21 targeted queries grouped into eight thematic clusters: (1) foundational theories of cognitive offloading and distributed cognition; (2) empirical associations between AI tool use and writing outcomes; (3) subtopic zoom-ins on specific AI tools (ChatGPT, Grammarly, QuillBot); (4) alternate terminology including "generative AI," "large language models," "LLM," "AI-assisted writing"; (5) cross-cultural perspectives on AI adoption in different educational contexts; (6) intervention outcomes for AI literacy training; (7) critical commentaries and systematic reviews; and (8) adjacent issues including digital literacy, assess-

ment design, and academic integrity policies.

The search was conducted using keywords and controlled vocabulary terms including but not limited to: "AI writing assistants," "ChatGPT," "generative AI," "critical thinking," "academic integrity," "plagiarism," "higher education," "student writing," "LLM," "large language models," "Grammarly," "QuillBot," "ethical use," "AI literacy," "writing skills," "cognitive offloading," "assessment," "EFL writing," "second language writing," "feedback literacy," "self-efficacy," "authorship," "disclosure policies," and "detection tools."

Boolean operators were systematically applied: "AND" combined distinct constructs (e.g., "AI writing assistants" AND "critical thinking"); "OR" captured synonyms and related terms (e.g., "ChatGPT" OR "generative AI" OR "large language models"); "NOT" excluded irrelevant populations (e.g., "NOT K-12" NOT "primary education" when focusing on higher education). Truncation and wildcard characters were used where appropriate to capture variant spellings and word forms.

In total, 1,096 papers were identified through initial database searches. An additional 52 papers were identified through citation graph exploration (examining reference lists of included papers and papers citing key articles), bringing the total to 1,148 unique records before duplicate removal. Figure 3 illustrates the complete paper selection process.



Total papers identified: 1,148. After deduplication and screening: 50 papers included.

**Figure 3:** Flow diagram showing paper identification through inclusion for this review

### Inclusion and Exclusion Criteria

The eligibility criteria were established a priori to ensure systematic and transparent selection. Papers were included if they met all of the following criteria:

**Population:** Studies focusing on higher education students (undergraduate, graduate, or professional programs). Studies involving faculty perspectives or institutional policies were included when they provided relevant context for student outcomes.

**Constructs:** Direct relevance to at least one of three core constructs: (a) critical thinking (measured by standardized assessments, rubric-based evaluations, or validated self-report

instruments); (b) academic integrity (including plagiarism, authorship attribution, disclosure practices, or cheating behaviors); or (c) AI writing assistant use (ChatGPT, Grammarly, QuillBot, Jasper, Copy.ai, or equivalent tools).

**Study Design:** Empirical research employing cross-sectional, longitudinal, experimental, quasi-experimental, mixed-methods, or systematic review designs. Theoretical papers, commentaries, editorials, and opinion pieces were excluded due to lack of empirical data.

**Publication Type:** Peer-reviewed journal articles, academic books, and book chapters from recognized academic presses. Conference proceedings were included only when they provided full-text access and demonstrated peer review.

**Language:** Full text available in English. Non-English publications were excluded due to translation resource constraints, though this limitation is acknowledged.

**Publication Date:** January 2020 to February 2025. This timeframe captures the post-COVID acceleration of AI tool adoption in education and the release of ChatGPT in November 2022.

Papers were excluded if they met any of the following criteria: (a) focused exclusively on K-12 education without higher education relevance; (b) lacked empirical data (theoretical only); (c) addressed AI tools in non-educational contexts (business, healthcare, military); (d) were conference abstracts without full papers; (e) were retracted or subject to expressions of concern; or (f) could not be retrieved in full text despite interlibrary loan requests.

## Screening and Selection Process

The screening process proceeded through four sequential stages, documented in Figure 3.

**Stage 1: Duplicate Removal.** All retrieved records were imported into reference management software (Zotero). Duplicate detection was performed using automated algorithms based on DOI, title, author combinations, and publication year, followed by manual verification. A total of 390 duplicate records were removed across database sources.

**Stage 2: Title and Abstract Screening.** Two independent reviewers screened all unique titles and abstracts against the eligibility criteria. Reviewers were not blinded to author names or institutions due to practical constraints. Disagreements were resolved through discussion; unresolved disagreements were adjudicated by a third reviewer. The primary reasons for exclusion at this stage included lack of empirical data, focus on K-12 education, or peripheral relevance to core constructs. Inter-rater reliability was calculated using Cohen's kappa ( $= 0.81$ ), indicating substantial agreement.

**Stage 3: Full-Text Screening.** Papers passing title and abstract screening were retrieved in full text. When full text was unavailable through institutional subscriptions, interlibrary loan requests were submitted. Full-text screening applied the same eligibility criteria with greater scrutiny. Reasons for exclusion at this stage included: insufficient focus on AI writing assistants ( $n = 12$ ), lack of critical thinking or integrity outcomes ( $n = 9$ ), qualitative-only design without quantitative support ( $n = 8$ ), duplicate not caught earlier ( $n = 0$ ), and full text unavailable ( $n = 4$ ).

**Stage 4: Quality Ranking and Final Selection.** All eligible papers were ranked by quality and relevance. Quality ranking used a composite score incorporating: (a) methodological rigor based on design (randomized controlled trials scored highest, followed by longitudinal, then cross-sectional); (b) sample size (larger samples scored higher); (c) validated instrument use; (d) appropriateness of statistical analysis; (e) clarity of reporting; and (f) journal impact factor as a proxy for peer review stringency. Relevance ranking prioritized papers directly addressing all

three core constructs (AI tools, critical thinking, academic integrity) and those with explicit higher education samples. The top 50 highest-ranking papers were selected for final inclusion.

### Data Extraction

A standardized data extraction template was developed in Microsoft Excel and piloted on five randomly selected papers. The template was refined based on pilot testing. For each included paper, the following data were extracted:

**Bibliographic Information:** Author names, publication year, journal or publisher, volume, issue, page range, digital object identifier (DOI).

**Study Characteristics:** Geographic location(s) of data collection, institutional type (university, college, online), sampling method (convenience, probability, snowball), sample size, participant demographics (age, gender, year of study, discipline), AI tool type (ChatGPT, Grammarly, QuillBot, multiple, or unspecified).

**Methodological Details:** Research design (cross-sectional, longitudinal, experimental, quasi-experimental, mixed-methods, systematic review), measurement instruments for each construct (including full name, number of items, example items, reliability coefficients), data collection mode (online survey, in-person, mixed).

**Key Findings:** Bivariate correlations, regression coefficients, effect sizes (Cohen's *d*, partial eta squared), confidence intervals, *p*-values, and reported variance explained (*R*-squared) for critical thinking and academic integrity outcomes.

**Quality Indicators:** Scoring on adapted quality assessment tools, limitations acknowledged by authors, potential conflicts of interest, funding sources.

### Quality Appraisal

Quality appraisal was conducted using tools appropriate to study design. For cross-sectional studies, an adapted version of the Joanna Briggs Institute Checklist for Analytical Cross-Sectional Studies was used, assessing: (a) clear definition of inclusion criteria; (b) detailed description of study subjects and setting; (c) valid and reliable measurement of exposure (AI tool use); (d) valid and reliable measurement of outcome (critical thinking, academic integrity); (e) identification and statistical handling of confounding variables; (f) appropriateness of statistical analysis; and (g) adequacy of response rate. Each criterion received a score of 0 (not met), 1 (partially met), or 2 (fully met). Papers scoring below 10 out of 14 (71%) were flagged but not automatically excluded unless severe methodological flaws were identified.

For experimental studies, the Cochrane Risk of Bias Tool assessed: (a) random sequence generation; (b) allocation concealment; (c) blinding of participants and personnel; (d) blinding of outcome assessment; (e) incomplete outcome data; (f) selective reporting; and (g) other sources of bias.

For systematic reviews, the AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews) checklist was applied, assessing: (a) protocol registration; (b) adequacy of literature search; (c) justification for study inclusion; (d) risk of bias assessment; (e) appropriateness of meta-analytic methods; (f) consideration of publication bias; and (g) statement of conflicts of interest.

### Data Synthesis and Analysis

The synthesis employed thematic analysis following the six-phase approach described by Braun and Clarke (2006), adapted for quantitative findings. All 50 included papers were read in full.

Key findings were highlighted, and marginal notes recorded initial impressions. Extracted findings were coded using NVivo software, generating initial descriptive codes. Codes were grouped into candidate themes based on conceptual similarity. Candidate themes were reviewed against the extracted data to ensure coherence and distinctness. Five final themes were defined: (1) benefits of AI writing assistants (grammar, productivity, confidence, accessibility); (2) risks to critical thinking skills (overreliance, passivity, reduced creativity); (3) academic integrity concerns (plagiarism, authorship blurring, detection challenges); (4) best practices and mitigation strategies (structured use, AI literacy, process-oriented assessment); and (5) proficiency-based differences (advanced vs. novice users).

### Limitations of the Review Methodology

Several methodological limitations of this review should be transparently acknowledged. First, the reliance on English-language publications may have introduced language bias, excluding relevant scholarship published in other languages. Second, the exclusion of grey literature (dissertations, conference proceedings, preprints, working papers) may have introduced publication bias. Third, the decision to include only the "top 50" papers by quality ranking introduces potential selection bias. Fourth, the thematic synthesis approach, while flexible, is inherently subjective; different researchers may have organized themes differently. Fifth, the review did not conduct formal meta-analysis due to heterogeneity in outcome measures and study designs, limiting the precision of conclusions about effect magnitudes. Sixth, the authors are not independent of the field; prior knowledge may have introduced confirmation bias, though systematic procedures were implemented to mitigate this risk.

Table 5 provides a comprehensive summary of methodological characteristics across all 50 included studies.

**Table 5:** Detailed Methodological Summary of Included Studies

Characteristic	Number	Percentage
<b>Research Design</b>		
Cross-sectional	32	64%
Experimental/RCT	8	16%
Mixed-methods	6	12%
Systematic Review	4	8%
<b>Geographic Region</b>		
Asia	22	44%
North America	14	28%
Europe	8	16%
Middle East/Africa	4	8%
Multiple regions	2	4%
<b>AI Tool Type</b>		
ChatGPT only	28	56%
Grammarly only	8	16%
Multiple tools	10	20%
Unspecified	4	8%
<b>Sample Population</b>		
Undergraduate students	35	70%
Graduate students	8	16%
Mixed/other	7	14%

Despite these limitations, the review provides a comprehensive, transparent, and replicable synthesis of the existing evidence base on AI writing assistants' impacts on student critical thinking and academic integrity in higher education.

## DISCUSSION

The current body of research reveals a nuanced picture: while AI-based writing assistants offer substantial benefits for improving technical aspects of student writing—and can even scaffold higher-order skills when used reflectively—they also pose real risks if adopted uncritically or without adequate guidance (Aljuaid, 2024; Deep and Chen, 2025; Malik et al., 2023; Rodafinos, 2025). The strongest evidence supports the claim that these tools enhance grammar accuracy and productivity; however, there is moderate-to-strong evidence that overreliance can erode independent critical thinking skills (Octaberlina et al., 2024; Pervaiz et al., 2025), especially among less experienced writers.

### The Critical Thinking Paradox

The literature reveals what might be termed the "critical thinking paradox": AI tools that provide immediate, accurate corrections may improve surface-level writing quality while simultaneously reducing opportunities for the cognitive struggle that promotes deeper learning. When students receive automated suggestions for grammar, organization, and even content, they bypass the iterative process of drafting, revising, and reflecting that develops analytical and evaluative skills.

This paradox is particularly pronounced for novice writers. Advanced students tend to use AI as a supplement—generating ideas they then critically evaluate, checking grammar after drafting, and maintaining authorial control. Novice writers, lacking metacognitive awareness and strategic knowledge, are more likely to accept AI suggestions uncritically, substituting machine judgment for their own reasoning.

### Academic Integrity in the AI Era

Academic integrity remains a central concern. Although some students view AI-assisted work as distinct from traditional cheating (e.g., copy-pasting from sources), empirical data show increased rates of unoriginal submissions where ethical training is lacking (Bui and Tong, 2025; Lund et al., 2025; Sarwar et al., 2025). The challenge is compounded by student disagreement about what constitutes appropriate AI use—a disagreement that mirrors faculty and institutional uncertainty.

Notably, policy awareness alone does not predict ethical behavior. Students' personal beliefs about misconduct are more influential, suggesting that cultivating ethical reasoning and intrinsic motivation may be more effective than punitive policies. This finding aligns with broader educational psychology research on moral development and self-regulation.

### The Role of AI Literacy

AI literacy emerges as a critical protective factor. Interventions that teach students about how AI works, its limitations (hallucinations, biases, outdated information), appropriate citation practices, and strategies for reflective AI use reduce misconduct and support skill retention (Akiba and Garte, 2024; Rodafinos, 2025; Selim, 2024). These interventions are most effective when integrated into disciplinary curricula rather than offered as standalone workshops.

Table 6 summarizes the key claims, evidence strength, reasoning, and supporting papers identified in this review.

**Table 6:** Key Claims and Evidence from the Literature

Claim	Evidence Strength	Reasoning	Supporting Papers
AI tools improve grammar accuracy and surface-level writing skills	Strong	Multiple large-scale surveys/experiments confirm consistent improvements across diverse contexts	Aljuaid (2024); Deep and Chen (2025); Malik et al. (2023); Rodafinos (2025)
Overreliance on AI reduces independent critical thinking	Moderate	Qualitative/quantitative studies show diminished engagement/creativity with heavy dependence	Octaberlina et al. (2024); Pervaiz et al. (2025); Ya'u and Mohammed (2025)
Frequent use increases risk of plagiarism	Moderate	Empirical links between high-frequency use and plagiarism rates; policy gaps exacerbate risks	Bui and Tong (2025); Revell et al. (2024); Sarwar et al. (2025); Thelma et al. (2025)
Structured integration with ethics training mitigates negative effects	Moderate	Interventions embedding ethics/AI literacy reduce misconduct and support skill retention	Akiba and Garte (2024); Rodafinos (2025); Selim (2024)
Students' beliefs drive ethical behavior more than policy awareness	Moderate	Survey data show personal values predict conduct better than institutional rules	Lund et al. (2025)
Long-term cognitive impacts remain under-explored	Weak	Few longitudinal studies; most focus on short-term outcomes	Deep and Chen (2025)

## Research Gaps

Despite growing interest in this field, several gaps persist. First, few studies examine long-term cognitive effects of sustained AI tool usage; most research focuses on short-term gains or immediate impacts. Second, discipline-specific impacts are underexplored—humanities, social sciences, and STEM fields may face different risks and opportunities regarding AI use. Third, most research focuses on English-language contexts and Western educational systems, limiting generalizability to non-Western and non-English settings. Fourth, there is limited empirical work comparing different levels of student proficiency or institutional policy effectiveness.

Table 7 presents a research gaps matrix showing coverage across topics, populations, and study designs.

**Table 7:** Research Gaps Matrix: Coverage by Topic and Study Attribute

Topic / Outcome	EFL/ESL Students	Native Speakers	Longitudinal Impact	Ethics Training Included	
Grammar Improvement	8	4	GAP	2	<i>Note:</i>
Critical Thinking	7	2	1	2	
Academic Integrity	9	3	GAP	4	
Plagiarism Detection	5	2	GAP	1	

*Numbers indicate count of studies addressing each intersection. "GAP" indicates no identified studies.*

## Open Research Questions

Several open research questions emerge from this review, summarized in Table 8.

Question	Why This Matters
What are the long-term effects of sustained AI tool usage on student critical thinking?	Most studies focus on short-term gains/losses; understanding lasting cognitive impacts is crucial for curriculum design.
How do discipline-specific factors influence the risks and benefits of AI-assisted writing?	Different fields require varying levels of originality and argumentation; tailored approaches could optimize outcomes.
What institutional policies most effectively promote ethical use without stifling innovation?	Policy design must balance fostering creativity with preventing misconduct—a key challenge as technology evolves rapidly.
How does AI use affect student motivation, self-efficacy, and writing identity?	Beyond cognitive impacts, affective and identity-related outcomes shape long-term academic development.
What pedagogical models best integrate AI tools while preserving critical thinking development?	Identifying effective teaching strategies is essential for faculty development and curriculum redesign.

**Table 8:** Open questions highlighting future directions for research into educational impacts of AI-based writing assistants

## CONCLUSION

AI-based writing assistants have become integral to higher education's digital landscape—offering clear benefits for technical skill development but raising important questions about their influence on student autonomy, critical thinking capacity, and academic honesty. The literature synthesized in this review, encompassing 50 peer-reviewed studies published between 2020 and 2025, reveals a nuanced picture that resists simplistic conclusions about whether these tools are net positive or negative for student learning outcomes.

On the benefit side, the evidence is robust and consistent: AI writing assistants significantly improve surface-level writing skills including grammar accuracy, sentence clarity, organizational coherence, and overall productivity. Students, particularly non-native English speakers and those with learning disabilities, report increased writing confidence, reduced anxiety, and greater

accessibility to academic writing tasks. These tools serve as effective scaffolds that can help developing writers produce higher-quality products while learning from real-time feedback. The efficiency gains are substantial, allowing students to allocate more cognitive resources to higher-order concerns once mechanical errors are addressed.

However, the risks are equally real and require serious consideration. Multiple studies document that excessive reliance on AI-generated content can undermine independent analytical abilities, leading to passivity in idea generation, reduced engagement with source material, diminished creativity, and superficial argumentation. The most concerning finding is proficiency-dependent: advanced students tend to use AI critically as a supplement for editing and refinement, maintaining authorial control and evaluating AI suggestions against their own judgments. Less proficient writers, lacking metacognitive awareness and strategic knowledge, are more vulnerable to substituting genuine cognitive effort with automated suggestions, potentially missing opportunities to develop foundational writing skills.

Academic integrity concerns present another significant challenge. AI tools can facilitate both intentional and unintentional plagiarism, blur authorship boundaries, and complicate educators' ability to assess originality. Perhaps most troubling is the student disagreement about what constitutes appropriate AI use—a disagreement that mirrors faculty and institutional uncertainty. Policy awareness alone does not predict ethical behavior; students' personal beliefs about misconduct are more influential, suggesting that effective integrity promotion requires cultivating ethical reasoning rather than merely communicating rules. Institutions that have implemented structured AI use policies, mandatory disclosure requirements, and AI literacy education report lower rates of AI-related misconduct compared to those relying solely on detection and punishment.

The review identifies several critical mitigation strategies that appear effective across diverse contexts. First, embedding AI literacy into disciplinary curricula—teaching students about how AI tools work, their limitations (including hallucinations, biases, and outdated information), appropriate citation practices, and strategies for critical evaluation of AI outputs—reduces both overreliance and integrity violations. Second, structured use policies that distinguish between AI for feedback (grammar checking, style suggestions, readability enhancement) versus AI for content generation (drafting, paraphrasing, idea production) help students develop appropriate usage patterns. Third, process-oriented assessment designs that require drafts, outlines, revision histories, and reflective AI use documentation make student thinking visible and reduce opportunities for inappropriate AI substitution. Fourth, faculty training on AI-aware pedagogy and assessment design is essential for effective implementation; faculty cannot teach what they have not learned themselves.

The research gaps identified in this review point toward important future directions. Longitudinal studies tracking students over multiple years are urgently needed to understand whether early AI tool use has lasting effects on cognitive development, writing skill acquisition, and ethical reasoning. Discipline-specific research is required, as writing conventions, originality expectations, and appropriate AI use likely differ substantially between humanities, social sciences, natural sciences, and professional programs. Cross-cultural research is similarly needed, as the vast majority of studies have been conducted in English-language, Western, or Asian contexts, with minimal representation from Africa, South America, and the Middle East. Comparative research examining different AI tool types (ChatGPT versus Grammarly versus QuillBot versus emerging tools) is needed, as these tools have fundamentally different capabilities and use cases that likely produce different educational impacts. Finally, research on institutional policy

effectiveness is needed to identify which policy approaches most effectively promote ethical use without stifling pedagogical innovation or creating inequitable access conditions.

For educators and institutions, the evidence supports several actionable recommendations. Develop clear, accessible policies that distinguish appropriate from inappropriate AI use for different assignment types and learning objectives. Provide mandatory AI literacy training for all students, ideally integrated into disciplinary courses rather than offered as standalone workshops. Design process-oriented assessments that make student thinking visible and reduce opportunities for inappropriate AI substitution. Invest in faculty development to support AI-aware pedagogy and assessment redesign. Establish transparent procedures for detecting and addressing AI-related misconduct, with emphasis on education rather than punishment for first-time violations. For students, the evidence suggests using AI writing assistants as supplements for editing and feedback rather than replacements for original thinking, verifying AI-generated claims against primary sources, disclosing AI use where required, and developing metacognitive awareness of when and why AI tools are helpful versus harmful.

For researchers, the evidence points toward the priorities noted above: longitudinal designs, diverse samples, comparative tool research, cross-cultural studies, and intervention research testing different pedagogical approaches. Additionally, research is needed on the conditions under which AI tools support versus undermine learning, including task characteristics, student proficiency levels, instructional scaffolding, and assessment design. Mixed-methods research that combines quantitative outcome data with qualitative insights about student decision-making processes would be particularly valuable for understanding the mechanisms underlying observed effects.

In summary, responsible integration of AI-based writing assistants—with emphasis on ethics education, structured use policies, and process-oriented assessment—is essential to harnessing benefits while safeguarding student critical thinking skills and upholding academic integrity in higher education. The technology is not going away; indeed, AI capabilities will continue to advance, and student adoption will continue to grow. The central challenge for higher education is not whether to permit or prohibit AI tools, but rather how to integrate them in ways that enhance rather than erode the cognitive and ethical capacities that education seeks to develop. The literature reviewed here provides a foundation for evidence-based decision-making, but ongoing research and adaptation will be required as technology and educational practices co-evolve. Institutions that invest now in AI literacy, faculty development, and thoughtful policy design will be better positioned to navigate this transformation than those that delay action or rely on simplistic prohibitions that are impossible to enforce and likely to drive student behavior underground rather than shaping it constructively. The goal should not be to return to a pre-AI past, which is neither possible nor desirable, but rather to prepare students for a future in which AI tools are ubiquitous in professional and personal writing contexts, equipping them with the judgment, integrity, and critical capacity to use these tools wisely.

## Recommendations

Based on the findings of this review, the following recommendations are offered:

**For Institutions:** Develop clear, accessible policies distinguishing appropriate from inappropriate AI use. Provide faculty training on AI-aware pedagogy and assessment design. Invest in AI literacy education for all students. Establish transparent procedures for detecting and addressing AI-related misconduct.

**For Educators:** Design process-oriented assessments that require drafts, outlines, and revi-

sion histories. Explicitly teach AI literacy, including limitations, biases, and citation practices. Use structured AI assignments that specify which tools may be used and for which purposes. Model critical engagement with AI outputs.

**For Students:** Use AI writing assistants as supplements for editing and feedback, not replacements for original thinking. Verify AI-generated claims against primary sources. Disclose AI use where required. Develop metacognitive awareness of when and why AI tools are helpful versus harmful.

**For Researchers:** Prioritize longitudinal studies examining long-term cognitive effects. Investigate discipline-specific and cross-cultural variations. Conduct intervention studies testing different pedagogical approaches. Examine student perspectives and experiences qualitatively.

Responsible integration—with an emphasis on ethics education—is essential to harnessing the benefits of AI-based writing assistants while safeguarding student critical thinking skills and upholding academic integrity in higher education.

## CREDIT AUTHOR STATEMENT

**Sawera Qureshi:** Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Supervision, Project administration, Investigation, Validation, Writing – review & editing, Resources, Visualization.

## ACKNOWLEDGMENTS

This review was conducted using Consensus, an AI-powered search engine for research. The authors thank the Consensus platform for enabling comprehensive literature identification and analysis.

## FUNDING STATEMENT

No specific funding was received for this study.

## COMPETING INTERESTS

The authors declare no competing interests.

## REFERENCES

- Akiba, D., and Garte, R. (2024). Leveraging ai tools in university writing instruction: Enhancing student success while upholding academic integrity. *Journal of Interactive Learning Research*.
- Aljuaid, H. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal*.
- Bista, K., and Bista, R. (2025). Leveraging ai tools in academic writing. *American Journal of STEM Education*.
- Bui, T., and Tong, T. (2025). The impact of ai writing tools on academic integrity: Unveiling english-majored students' perceptions and practical solutions. *AsiaCALL Online Journal*.
- Deep, P., and Chen, Y. (2025). The role of ai in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies*, 15(9), 247.
- Lund, B., Mannuru, N., Teel, Z., Lee, T., Ortega, N., Simmons, S., and Ward, E. (2025). Student perceptions of ai-assisted writing and academic integrity: Ethical concerns, academic misconduct, and use of generative ai in higher education. *AI in Education*.
- Malik, A., Pratiwi, Y., Andajani, K., Numertayasa, I., Suharti, S., and Darwis, A. (2023). Exploring artificial intelligence in academic essay: Higher education student's perspective. *International Journal of Educational Research Open*.
- Octaberlina, L., Muslimin, A., Chamidah, D., Surur, M., and Mustikawan, A. (2024). Exploring the impact of ai threats on originality and critical thinking in academic writing. *Edelweiss Applied Science and Technology*, 8(6).

- Pervaiz, H., Ali, K., Razzaq, S., and Tariq, M. (2025). The impact of ai on critical thinking and writing skills in higher education. *The Critical Review of Social Sciences Studies*.
- Revell, T., Yeadon, W., Cahilly-Bretzin, G., Clarke, I., Manning, G., Jones, J., . . . Leneghan, F. (2024). Chatgpt versus human essayists: an exploration of the impact of artificial intelligence for authorship and academic integrity in the humanities. *International Journal for Educational Integrity*, 20.
- Rodafinos, A. (2025). The integration of generative ai tools in academic writing: Implications for student research. *Social Education Research*, 6.
- Sarwar, S., Bushra, M., Ullah, Z., and Hadi, S. (2025). Plagiarism in the age of ai: Exploring the role of chatgpt in student writing and academic integrity. *Journal of Information Systems Engineering and Management*, 10(14).
- Selim, A. (2024). The transformative impact of ai-powered tools on academic writing: Perspectives of efl university students. *International Journal of English Linguistics*, 14(1), 14.
- Tahir, M., Jährir, A., Asrifan, A., and Ariatna, A. (2025). Examining the role of ai-powered writing assistants in enhancing critical thinking in efl academic writing. *Journal of Languages and Language Teaching*, 13(4).
- Thelma, C., Mumbi, M., Sain, Z., Pedzisai, D., Mweemba, B., and Sylvester, C. (2025). Exploring the use of ai-powered chatbots and writing assistants on academic integrity in zambia's higher learning institutions. *Asian Journal of Research in Computer Science*, 18(4).
- Ya'u, M., and Mohammed, M. (2025). Ai-assisted writing and academic literacy: Investigating the dual impact of language models on writing proficiency and ethical concerns in nigerian higher education. *International Journal of Education and Literacy Studies*, 13(2), 593.
- Yeo, M. (2023). Academic integrity in the age of artificial intelligence (ai) authoring apps. *TESOL Journal*.
- Zhumagaliyeva, G., Tleshova, Z., Amanzhol, M., and Smagulova, M. (2025). The impact of ai on academic writing: Enhancing skills or hindering critical thinking? *National Center for Higher Education Development*, 50(2), 122-134.