Use of Artificial Intelligence-Based Writing (AIW) Tools in Higher Education: Opportunities, Challenges, and Future Directions for Sustainable Academic Practices

Authors: Lilliati Ismail[1]*, Sahar M. Taresh[1], Nooreen Noordin[1], Ilyana Jalaluddin[2]
Affiliation: [1]Faculty of Educational Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2]Faculty of Modern Languages and Communication, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
*Corresponding author: liliati@upm.edu.my

Received: 31/12/2025 | First Revision: 14/1/2026 | Second Revision: 15/4/2026 | Accepted: 15/5/2026

APA citation for this article:

ABSTRACT

The use of AI-based writing (AIW) tools in academic writing has become increasingly common in higher education. At the same time, their implications for writing pedagogy and academic integrity remain widely debated. This systematic review synthesises research published between January 2020 and August 2025 on the benefits and challenges of AIW tools in higher education, with particular attention to pedagogy, ethics, inclusivity, and institutional policies. Peer-reviewed studies were identified through searches of major databases, including Scopus, Web of Science (Core Collection), and Embase. Studies were screened and selected using predefined inclusion and exclusion criteria. Due to variation in study designs and outcome measures, the findings were synthesised using a narrative thematic approach. Methodological quality and potential sources of bias were assessed using the Mixed Methods Appraisal Tool (MMAT, 2018), and study quality was considered when interpreting the results. A total of 27 studies were included in the review. The findings show that AIW tools can support academic writing by providing immediate feedback, assisting non-native English speakers, and improving writing efficiency. However, the review also identifies significant concerns related to academic integrity, authorship attribution, equitable access, and possible negative effects on higher-order thinking skills. The existing evidence base is limited by short-term study designs, context-specific findings, and a lack of longitudinal and policy-focused evaluations. Overall, this review highlights the need for responsible, fair, and sustainable integration of AIW tools in higher education. Such integration should balance technological opportunities with core educational values and contribute to the goal of Quality Education (SDG 4). This systematic review was registered with the Open Science Framework (OSF) (10.17605/OSF.IO/F8P6C)

Keywords: Artificial intelligence (AI), higher education, academic writing, academic integrity, AI-based writing (AIW) tools.

1. Introduction

It is undeniable that artificial intelligence (AI) has significantly changed the educational landscape at all levels including in higher education. This includes the use of AI-based writing (AIW) tools in academic writing. AIW tools include Grammarly, ChatGPT, WriteFull, and QuillBot. The use of these tools have significantly transformed the manner in which students engage with their academic writing tasks (Imran & Norah, 2023; Moorhouse, Wan, Wu, Wu, & Ho, 2025).

Academic writing is an integral part of the higher education experience. There are rules and conventions in academic writing which students must master. These include presenting coherent and well-substantiated discussions and arguments, using appropriate academic English and adhering to citation regulations (Hind, 2024). Synthesizing, analyzing and evaluating a huge amount of past literature for the literature review, identifying gaps from past literature and discussing how the present study addresses these gaps can be a highly demanding task for students. In addition, they would also need to identify and discuss theories in relation to their study. All these tasks are cognitively demanding and time consuming. For many students, the challenges in academic writing are further compounded by lack of English language proficiency. Thus, AIW tools are viewed as a potential solution to overcoming these challenges.

However, the opportunities afforded by AIW tools come together with great ethical concerns. The ease and efficiency in which they generate sophisticated and linguistically coherent content, has triggered widespread concern of academic misconduct and decline in critical thinking and writing mastery among students. The lack of intellectual contribution, originality and critical engagement in the writing is particularly concerning as the assessment and grade the student receives might not be a true reflection of the students' ability and content mastery (e.g., Hazem, John, & Mike, 2023; Johnston, 2024). Critics also warn that excessive reliance on AI may undermine the development of critical thinking, analytical reasoning, and independent writing skills competencies that are central to the purpose of higher education (Gallent, Zapata, & Ortego, 2023; Kleopatra, 2024).

Higher education institutions globally have started to respond to these downsides of AIW tools by developing suitable guidelines and policies to ensure responsible use of AIW tools. However, these institutions face challenges in determining the permissible level for using AI, and the effectiveness and credibility of AI-detection tools. Often AI-detection tools cannot make definite determination of AI use and can only suggest likelihood of AI use. The rapid and dynamic revolution of AI also makes developing guidelines a challenge for higher education institutions. As AI continues to grow in robustness and sophistication, guidelines can quickly become obsolete. In addition to pedagogical and ethical considerations, the sustainable integration of AIW tools, aligned with the goal of Quality Education (SDG 4) has become an increasingly important concern for higher education institutions.

Recent systematic reviews have also started studying the use of artificial intelligence writing (AIW) tools in higher education; although the scope is still comparatively limited and divided. For example, recent reviews have concentrated more on general AI uses in education or initial evidence about the tools like ChatGPT, and they tend to be more tuned to the short-term pedagogical effects and ethical considerations without considering the bigger institutional picture (e.g., Kasneci et al., 2023; Zawacki-Richter et al., 2019). Although these studies are extremely valuable as frameworks, they are limited in the synthesis that would integrate governance, inclusiveness, and sustainability as an analytical framework. In addition, the reviews before 2022 can be dated by the time the new generation AI technologies got out of the rapid growth phase, which restricts their applicability to the present educational setting. On the contrary, the current systematic review builds on the current body of knowledge in three major aspects. First, it covers a period that is more recent and urgent (2020-2025), both the digital transformation driven by the pandemic and the advent of more powerful generative AI systems. Second, it uses a multidimensional prism that clearly incorporates the elements of pedagogical, ethical, governance, inclusivity, and sustainability, thus filling gaps in previous syntheses. Third, it improves methodological clarity by registering on Open Science Framework (OSF) and providing

extensive supporting material, which facilitates reproducibility and rigor in accordance with modern standards of systematic review.

To bring even more analytical focus on the review, the objectives are operationalized by a set of guiding sub-questions. First, in what ways do various categories of AI-based writing aids (i.e. grammar and style aids, paraphrasing aids, and generative AI systems (i.e. ChatGPT)) differ in terms of pedagogical value and risks? Second, how do the stated advantages and difficulties differ between the student groups, specifically between multilingual (L2) student learners and native (L1) speakers, and learners who have a low level of academic achievements and those who have a high level of academic achievements? Third, which institutional and contextual conditions mediate the effective and responsible use of AIW tools in higher education including assessment design, policy frameworks, and access conditions? These sub-questions enhance the overall review goals since they allow differentiated and theoretically informed synthesis of the literature especially regarding tool functionality, learner attributes and institutional governance.

2. Theoretical and Conceptual Framework

This systematic review is grounded in well-established theories of learning and writing. In particular, it draws on the sociocultural theory and self-regulated learning theory, both of which have long been used to explain how learners develop complex skills such as academic writing.

From a sociocultural perspective, learning takes place through mediated support that is gradually reduced as learners gain competence (Vygotsky, 1978; Wood, Bruner, & Ross, 1976). In writing research, scaffolding commonly refers to instructional supports such as feedback, modelling, prompts, and guided revision that help students move from novice to more independent writers. More recent research in educational technology has extended this idea by viewing digital tools as cognitive or pedagogical scaffolds that support different stages of the writing process (Graham & Harris, 2018; Ranalli, 2018).

Alongside this, self-regulated learning theory emphasises the active role of learners in planning, monitoring, and evaluating their own learning (Zimmerman, 2002). Academic writing is widely recognised as a self-regulated activity that involves drafting, revising, evaluating sources, and responding to feedback. Technologies that provide support during drafting and revision may therefore influence not only the quality of written texts but also how learners regulate their writing processes.

Building on these perspectives, this review adopts a conceptual framework that views AI-based writing (AIW) tools as scaffolding supports within the writing process. Rather than treating AIW tools as replacements for learning, the framework conceptualises them as supports that shape key learner processes, including drafting, revising, evaluating content, and reflecting on writing decisions. This view is supported by several empirical studies included in the review, which show that students commonly use AIW tools for early drafting, feedback, and revision support (e.g., Angelos et al., 2024; Oates & Johnson, 2025; King et al., 2024).

At the same time, drawing on work in academic integrity and educational ethics, the framework recognises that scaffolding can become counterproductive when it is poorly guided or misused (Bretag, 2016; Eaton, 2021). When AIW tools are used without clear pedagogical structure, students may bypass cognitive effort, leading to concerns related to authorship, plagiarism, and surface-level learning. These risks are documented in several of the reviewed studies (AlAfnan et al., 2023; Duah & McGivern, 2024).

The framework also incorporates institutional governance as a key moderating layer. Research on educational policy and assessment design shows that learning technologies do not operate in isolation; their effects are shaped by institutional policies, assessment practices, lecturer guidance, and access conditions (Bearman et al., 2022; Kővári, 2025). Consistent with this view, the reviewed studies suggest

that factors such as institutional policies, assessment redesign, staff training, and equity-focused infrastructure strongly influence whether AIW tools function as productive learning supports or as sources of ethical and educational risk.

Figure 1 presents the conceptual framework adopted in this review. The framework links AIW tool affordances to learner processes and learning outcomes, while highlighting the central role of institutional governance in shaping responsible, ethical, and sustainable use aligned with Quality Education (SDG 4).

Drawing on scaffolding theory and self-regulated learning, the framework positions AIW tools as pedagogical supports that influence proximal writing processes, including drafting, revision, evaluation, and reflection. These processes, in turn, shape more distal outcomes such as writing skills, critical thinking, and academic integrity. Institutional governance mechanisms such as policy development, assessment design, lecturer training, access, and sustainability considerations moderate these relationships and guide responsible AI integration. This framework informed the data extraction and thematic analysis by guiding the categorisation of pedagogical, ethical, inclusivity, and governance-related findings.

Finally, research on AIW tools has evolved rapidly since 2020, particularly following the introduction of large language models such as ChatGPT. By adopting a PRISMA-based systematic review approach, this study aims to provide educators, institutional leaders, and policymakers with a clear, evidence-based understanding of the opportunities and challenges that AIW tools present for higher education.

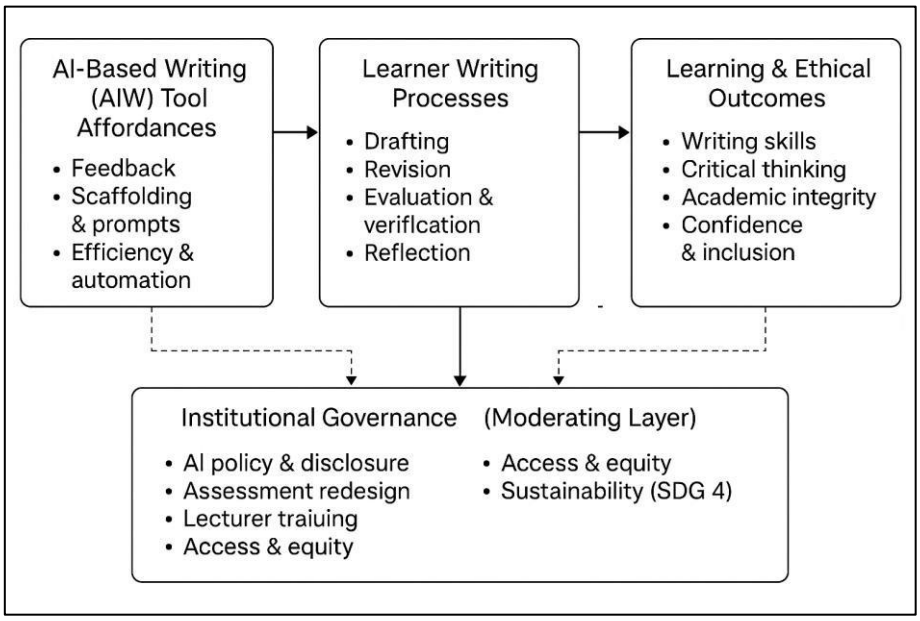

Figure 1. Processes, Learning Outcomes, and Institutional Governance

3. Methodology

3.1. Review Design

The current study was conducted in two phases. The first phase involved a systematic review, followed by a thematic analysis. The recommendations of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement were followed for conducting the systematic literature review. The following subsections detail the search strategies, eligibility criteria, data extraction, quality assessment, and analytical methods.

For the review, publications were collaboratively collected and evaluated by the first and second researchers using Google Drive, which facilitated the planning and management of the systematic literature review. This study was conducted using the systematic literature review methodology because it provides a comprehensive and clear overview of the literature written to date and identifies gaps and future research topics.

3.2. Review Process

The current review followed the five-step methodology for systematic literature reviews proposed by Denner and Tranfield (2009), which divides the review process into question formulation, study identification, study selection and evaluation, analysis and synthesis, and results reporting. The review protocol was retrospectively registered on the Open Science Framework (OSF) platform after the initial search and screening process was completed (OSF; DOI: 10.17605/OSF.IO/F8P6C). Such a retrospective registration was conducted to increase transparency and make sure that the review processes would be documented publicly. Notably, the research questions, inclusion criteria, search strategy, and data extraction procedures were not substantially modified once the screening process started. The registered protocol thus clearly shows the methodological approach applied in this study. Moreover, the OSF platform offers a record of the protocol under version control, which allows complete transparency about the moment of registration and the content. In addition, the scopes, eligibility criteria, and general analytical constructs of the study were not influenced by any minor procedural clarifications that were made during the review process.

The overall review flow also aligns with the PRISMA 2020 guidelines, and the main stages are summarized in Figure 2, with details of each stage provided.

Figure 2. Steps in Conducting the Systematic Literature review (adapted from Denyer and Tranfield, 2009)

3.2.1. Formulating the Questions

The first step in the systematic review process is formulating the questions. The review's focus was defined as the role of AI-powered writing tools in higher education. After several rounds of discussion among the researchers, four guiding questions were agreed upon to examine the following: (1) What AI-powered writing tools are used in university settings? (2) How do these tools affect academic writing and learning processes? (3) What challenges related to ethics, equity, and integrity are present? (4) What governance and sustainability issues have been discussed in the literature?

These questions were developed according to a population-concept-context logic, where higher education represents the population/context, and AI-powered writing tools are the central concept. These questions were used to guide and systematically develop the research strategy, eligibility criteria, and data extraction model structure, ensuring alignment between the review objectives, data collection, and objective analysis. Each research question was addressed through thematic grouping, with topics explicitly linked to guiding questions to ensure analytical transparency (see Supplementary File A).

3.2.2. Study Identification

The second step was locating studies, involved designing and implementing a comprehensive search strategy. We searched three electronic databases Scopus, Web of Science (Core Collection) (full), and Embase, because together they cover education, psychology, health, and interdisciplinary higher education research. The final searches were conducted between January and August 2025 (UTC).

A pilot search was first run in Scopus to test and refine keywords and controlled vocabulary. The final search strategy combined terms related to artificial intelligence and generative AI *(e.g., "artificial intelligence," "AI," "ChatGPT," "GPT-4," "generative AI," "AI writing tools"), academic writing (e.g., "academic writing," "writing," "composition"), and higher education (e.g., "higher education," "university," "college," "postsecondary"),* using Boolean operators AND and OR. Search limits were applied to restrict results to studies published between 2020 and 2025, written in English, and indexed as peer-reviewed articles or conference proceedings. Similar strategies were adapted for each database to account for differences in indexing and syntax. The complete search strategies for all databases, including exact Boolean strings and field tags, are provided in Supplementary File B, reported in accordance with PRISMA-S guidelines. All retrieved references were exported to EndNote for initial management and removal of duplicate records, and the reference lists of included studies were manually screened to identify additional relevant publications. No additional grey literature sources were searched.

3.2.3. Study selection and evaluation

The third step, study selection and evaluation, focused on screening the located records against predefined inclusion and exclusion criteria derived from the review questions. We included studies published in peer-reviewed journals or conference proceedings between January 2020 and August 2025, written in English, and situated in higher education contexts (undergraduate, postgraduate, or academic staff). To be eligible, studies had to examine the use of AI-based tools for academic writing, such as generative AI chatbots, grammar and style assistants, or paraphrasing tools, and report on at least one of the following aspects: pedagogical or learning outcomes, academic integrity or ethical issues, equity and accessibility, or institutional policies and governance. We excluded publications that focused on primary or secondary schooling, discussed AI in education without a clear link to writing, were opinion pieces, editorials, or news items without an empirical or clearly described methodological basis, were not available in full text, or were written in languages other than English.

After deduplication in EndNote, all remaining records were imported into Google drive (Excel) to facilitate independent and blinded screening by two reviewers. Titles and abstracts were first screened to remove clearly irrelevant studies; full texts were then retrieved for the remaining records and assessed against the same eligibility criteria in Google Drive (Excel). Disagreements at both stages were resolved through discussion and consensus. At this stage, we (first and corresponding authors) also evaluated the methodological quality of each included study using the Mixed Methods Appraisal Tool (MMAT, 2018). Quality appraisal was conducted independently by two reviewers, focusing on study design, clarity of research questions, sampling procedures, and appropriateness of analysis. Quality ratings were not used as exclusion criteria but were used to inform the interpretation and weighting of evidence during synthesis. A summary of MMAT assessments for all included studies is provided in Supplementary File C. The study selection process is illustrated in PRISMA Figure 3.

To limit the search strategy to peer-reviewed journal articles, methodological rigor, quality, and consistency of the evidence base were ensured. Grey literature sources (e.g. institutional reports, policy documents, conference proceedings, theses) were excluded, as such sources may differ in terms of methodological transparency and are not always indexed in databases like Scopus, Web of Science and Embase. The reason behind this choice is that the study is aimed at integrating empirically validated results into a regulated and comparable data set.

To enhance interpretability of the findings, the strength of evidence supporting each theme was qualitatively assessed based on the number, methodological quality, and consistency of the included studies. Evidence strength was categorized into three levels:

1. **Strong evidence**: supported by multiple studies (≥5) with consistent findings, including at least two studies of high methodological quality based on MMAT criteria

2. **Moderate evidence**: supported by a moderate number of studies (2–4), with generally consistent findings but some variation in quality or results
3. **Limited evidence**: supported by one study or a small number of studies with inconsistent findings or lower methodological quality

This approach allowed for a more comprehensive interpretation of the results and facilitated transparency in linking conclusions to the underlying evidence base. A detailed breakdown of exclusion reasons at both the title/abstract and full-text screening stages is provided in Table 1, and is fully consistent with the PRISMA flow diagram (Figure 3), ensuring transparency and reproducibility.

3.2.4. Analysis and Synthesis

In the fourth stage, analysis and synthesis, data from the included studies were extracted into a structured template. For each study, bibliographic information, country, research design, participants, type of AI-based writing (AIW) tool, educational context, and key findings related to pedagogical outcomes, academic integrity, inclusivity, and institutional governance were recorded.

A thematic synthesis approach was employed, combining deductive and inductive coding strategies. Initial coding categories were informed by the study objectives and conceptual framework, while additional themes emerged inductively from the data. The unit of analysis was the key findings reported in each included study. The extracted data were read repeatedly to identify recurring patterns.

Two reviewers independently conducted initial open coding using a structured coding framework (see Supplementary File E). Coding and data management were supported using Microsoft Excel. An initial calibration exercise was conducted on a subset of studies to align coding interpretations. Discrepancies between reviewers were resolved through iterative discussion and consensus.

With the help of an iterative process of comparison and refinement, the codes were clustered into descriptive sub-themes, and then grouped into four overarching analytical themes, namely pedagogical benefits, ethical and integrity challenges, inclusivity and accessibility, and institutional governance and sustainability. Instead of meta-analysis, narrative thematic synthesis was chosen because the included studies had varied study design, outcome measures, and reporting formats.

In line with the instructions of qualitative synthesis, there were no formal inter-coder reliability statistics (e.g., Cohens kappa). Nevertheless, the application of independent parallel coding as well as systematic consensus discussions is generally accepted in systematic reviews and guarantees the reliability and consistency of the analysis.

The conflicting results of the studies were not ruled out but analyzed in terms of the quality of the methods, the context of the studies, and the characteristics of the sample. This method allowed discovering convergent and divergent patterns. Analytical procedures were strongly related to the conceptual framework of the study, and thematic synthesis was both informed by the empirical patterns and theoretical constructs. The entire coding framework and the examples are detailed in Supplementary File D.

3.2.5. Reporting the Findings

In the final step of the review, the extracted data were organized according to the main themes that emerged from the analysis. These themes were clearly linked to the review questions and the theoretical framework to ensure consistency and coherence. Each theme presented its findings with direct reference to the studies from which the evidence originated, allowing readers to trace the results back to their original sources.

Descriptive statistics, summary tables, and conceptual diagrams were used to illustrate the characteristics of the included studies and to demonstrate the relationships between key opportunities,

challenges, and governance issues. The use of descriptive statistics was limited to summarizing study characteristics and general patterns in the literature; it was not used for inference or statistical testing.

In order to make sure that the theoretical framework would be a guide of analysis and not a justificatory role, we used theory-informed codes based on sociocultural and self-regulated learning views. The coding of studies was based on the features of scaffolding (e.g., its presence, compatibility with the learner ZPD, fading as the learner becomes more competent) and self-regulation indicators (e.g., goal-setting, monitoring, strategic use vs. dependency, evaluation of AI outputs). These codes based on the underpinning theories were used in conjunction with inductive codes. These were used to interpret the themes during synthesis to clarify why benefits and challenges arose and what could be identified as the discrepancies between tool use and theoretical principles.

The implications of these findings for sustainable and inclusive higher education are discussed in the following sections. Furthermore, key research gaps were identified, and practical, evidence-based recommendations were offered for educators, institutions, and policymakers.

4. Results

4.1 Study Selection

The selection of the studies was based on the PRISMA 2020 principles and is presented in Figure 3. The first search of the database occurred in Scopus, Web of Science, and Embase and resulted in n = 650 records. The elimination of duplicates resulted in n = 250.

After initial screening on the basis of the publication years, the type of document and its topic, n = 210 records were filtered out. The other n = 190 papers were then filtered through titles and abstracts leading to the removal of n = 145 records that were unrelated to AI-based writing aids in the higher education context and did not serve the purpose of the review.
N = 45 full-text articles underwent the evaluation of eligibility. Among these, n = 17 articles were excluded because of the inability to access the full texts and n = 1 study excluded because of the population criteria (i.e. not higher education). In the end, a total of 27 studies were included in the final analysis, as they fulfilled all inclusion criteria.

Table 1 presents a comprehensive analysis of the causes of exclusion at each phase and allows complete openness and repeatability of the selection of the study sample.

Table 1: Reasons for Study Exclusion

Stage	Reason for Exclusion	Number
Initial filtering	Year, document type, subject mismatch	210
Title/Abstract	Not AIW-related / not higher education	145
Full-text	Full text unavailable	17
Full-text	Wrong population	1

The literature search plan was formulated based on a combination of key words and Boolean operators, in order to get the studies associated with artificial intelligence writing aids in higher education. The keywords were the following: AI writing tools, automated writing help and the names of specific tools (e.g., ChatGPT), along with keywords connected to education (e.g., higher education, university, academic writing). In Scopus, Web of Science and Embase, searches were done using consistent Boolean logic (AND/OR), truncation and field restrictions where necessary.

The complete PRISMA 2020 flow diagram illustrating the study selection process is presented in **Figure 3**,

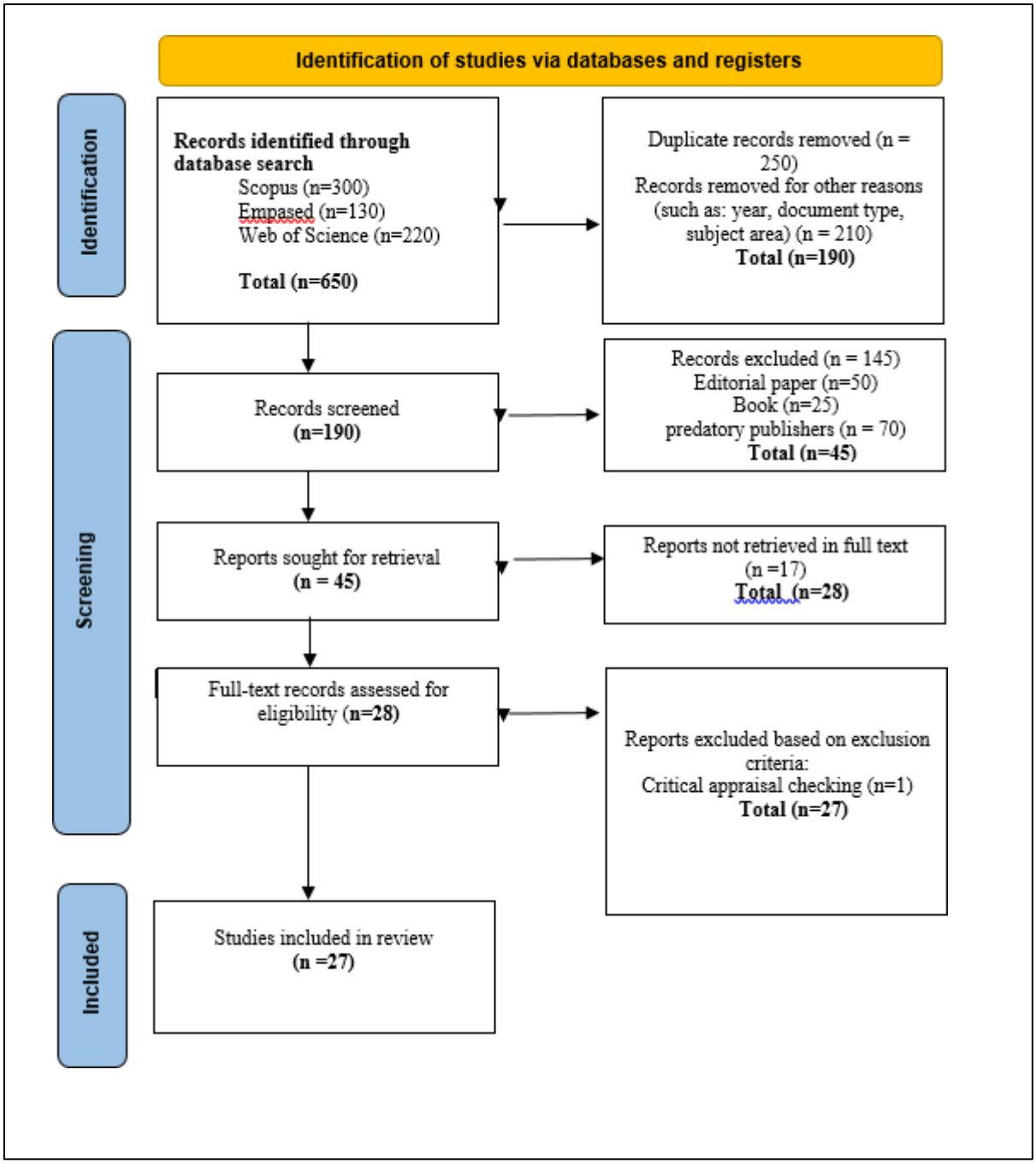

Figure 3. PRISMA Guideline

To enhance the transparency and enable the tracing of the synthesis, an evidence matrix that summarizes the nature of all the included studies is presented in Supplementary File C. The matrix specifies the setting of each study, research design, type of AIW tool used, contributions to the topic area, and quality appraisal scores, allowing the reader to trace the findings of each study to the synthesis.

An evidence matrix was created to show a systematic compilation of extracted data within studies categorized by study characteristics, type of tools, thematic coding, and quality of assessments (see Supplementary File D). Feedback that AIW tools gave in real-time was found to correlate with better quality of writing and revision. Findings from multiple past studies provide strong evidence to support this claim. Issues of academic honesty, such as fears of plagiarism and authorship, were also commonly reported in the literature (strong evidence). Conversely, the effect of AIW tools on critical thinking was inconsistent, and some studies depicted less cognitive engagement, and other studies indicated some

potential benefits when applied in a reflective manner (moderate evidence). Problems of accessibility and inclusivity were defined in a fewer number of studies, especially concerning the cost and infrastructure obstacles (limited evidence).

Table 2. Key Finding for Each Theme

Theme	Key Finding	Evidence Strength
Pedagogical Benefits	Improved writing quality	Strong
Ethical Challenges	Academic integrity concerns	Strong
Inclusivity	Access inequality	Limited

The classification of evidence strength is informed by the number of supporting studies and their methodological quality, as detailed in Supplementary File D.

4.2. Characteristics of Included Studies

4.2.1. Publication Year and Study Design

A total number of 27 studies were included. These studies represent a broad range of geographical regions, research methods, and disciplinary contexts. The studies used diverse research designs, including systematic and mapping reviews, survey-based studies, mixed-methods research, classroom-based interventions, as well as qualitative and quantitative approaches. Table 1 shows a summary of the studies, to illustrate the main types of evidence in the literature (e.g., Imran & Almusharraf, 2023; AlAfnan et al., 2023; Gustilo et al., 2024; Costa et al., 2024; Zhao et al., 2024; Kővári, 2025).

1

Table 3. Summary of Studies

Author(s) & Year	Study Type / Method	Sample / Context	AI Tools / Technologies Examined	Key Findings (Opportunities, Challenges & Sustainability)
Almashy et al. (2024)	Experimental comparative study (pre/post-test)	ESL/EFL learners divided into four groups (ChatGPT, Grammarly, Google Translate, control)	ChatGPT; Grammarly; Google Translate (CALL tools)	All AI tools significantly reduced writing errors, but tool-specific strengths were observed: ChatGPT improved verb forms and subject–verb agreement; Grammarly excelled in prepositions; Google Translate improved articles and capitalization. Findings support targeted, pedagogically guided AI use rather than general reliance, highlighting AI as a supportive tool for sustainable language learning.
Aqeel et al. (2025)	Narrative review / policy-oriented analysis	Higher education in Pakistan; applied and social sciences	ChatGPT; Grammarly; QuillBot; Turnitin	AI writing tools enhance grammar, content generation, paraphrasing, and research support, especially for under-resourced and non-native English-speaking students. Challenges include plagiarism, over-reliance, misinformation, bias, and weak policy frameworks. The study calls for ethical AI governance, institutional policies, AI literacy, and balanced human–AI collaboration to sustain academic integrity.
Aulia & Yamin (2024)	Systematic review (PRISMA 2020)	15 studies from multiple countries; EFL secondary & higher education	ChatGPT (writing assistance)	Identifies three major limitations of ChatGPT in EFL writing: (1) inability to support the entire writing process, (2) over-reliance and threats to originality, and (3) inaccurate or misleading feedback. While useful for idea generation and language support, ChatGPT requires strong teacher guidance and ethical oversight to ensure sustainable academic writing practices.
Bates, Cobo, Mariño & Wheeler (2020)	Editorial / conceptual synthesis (special issue introduction)	Higher education systems internationally	Artificial Intelligence (general educational AI applications)	Despite strong expectations, there is limited empirical evidence that AI has significantly improved teaching and learning outcomes in higher education. Most demonstrated impact lies in learning analytics and prediction, rather than pedagogy or writing quality. The paper cautions against technological hype, arguing for human- centred, augmentation-based AI use rather than replacement. Provides a critical foundation for understanding the gap between AI promises and sustainable academic realities.
Bin-Nashwan, Sadallah & Bouteraa (2023)	Quantitative empirical study (survey + SEM)	702 academics and researchers worldwide (ResearchGate & Academia.edu users)	ChatGPT	ChatGPT use is positively influenced by time-saving benefits, academic self-efficacy, self-esteem, perceived stress, and e-word of mouth. Academic integrity negatively affects adoption, yet also moderates usage behavior. Findings highlight the tension between

				productivity gains and integrity risks, stressing the need for clear ethical frameworks and institutional governance.
Bui & Tong (2025)	Mixed-methods empirical study (survey + semi-structured interviews; thematic analysis)	114 English-major seniors at a Vietnamese university	ChatGPT; Grammarly; QuillBot; AI writing tools	Students perceive AI writing tools as useful for grammar correction, idea generation, and writing efficiency, especially for non-native English speakers. However, major concerns include academic dishonesty, over-reliance, blurred authorship, and difficulty detecting AI-generated text. The study stresses the need for clear institutional guidelines, ethical training, and responsible AI use to preserve academic integrity.
Cheuquepán-Maldonado & Chourio-Acevedo (2025)	Quasi-experimental study; text-mining analysis; Mann–Whitney U tests	160 Spanish-speaking vocational education students in an online communication literacy course (Chile)	Generative AI embedded in an LMS discussion forum (AI-powered feedback system)	Use of generative AI–supported forums did not result in statistically significant improvements in textual readability compared to traditional forums (FHR $p = 0.233$; SPP $p = 0.262$). While AI assisted with idea generation and revision, it did not measurably enhance surface-level writing quality. The authors suggest that AIW benefits may lie beyond readability—such as coherence, cohesion, and higher-order writing skills—and emphasize the need for multi-dimensional assessment frameworks. Sustainable AIW integration requires pedagogically guided use and broader evaluation metrics, not reliance on superficial writing indicators alone.
Chalkiadakis et al. (2024)	Systematic review	Global higher education context; focus on students with disabilities and inclusive education	Artificial Intelligence (AI); Virtual Reality (VR); adaptive learning systems	AI-driven technologies support personalised learning, accessibility, and inclusion, particularly for students with disabilities. Challenges include high implementation costs, technical barriers, algorithmic bias, privacy concerns, and limited teacher readiness. Although not focused exclusively on writing tools, the review provides important sustainability and equity perspectives relevant to AI-assisted academic practices, highlighting the need for ethical governance and inclusive design.
Costa, Ntsobi & Mfolo (2024)	Systematic mapping review (research article; review-based)	Literature from 22 countries; higher education academic writing context (with focus on South Africa)	ChatGPT; Generative AI (GAI) writing tools	ChatGPT significantly enhances productivity, language support for non-native English speakers, and immediate feedback in academic writing. However, it presents serious challenges related to plagiarism, fabricated references, accuracy, over-reliance, and erosion of critical thinking. The study proposes the OTHA Framework (Openness, Transparency, Honesty, Accountability) to guide ethical and sustainable AI use. Recommends institutional

				policies, disclosure of AI use, ethical training, and monitoring to ensure responsible academic writing practices.
Chanpradit et al. (2024)	Qualitative descriptive case study; text analysis using Keck's (2014) and Nabhan et al.'s (2021) frameworks	30 Scopus-indexed research abstracts paraphrased by AI tools	QuillBot; Paraphrasing Tool (standard modes)	Both AI paraphrasing tools primarily rely on synonym substitution and sentence restructuring. QuillBot tends toward minimal to moderate revision, while Paraphrasing Tool shows more moderate revision, with limited instances of substantial rewriting. Neither tool consistently achieves deep conceptual transformation, indicating surface-level paraphrasing. The study warns that AI paraphrasing tools may reduce genuine paraphrasing skill development and pose academic integrity risks if overused, recommending teacher-guided use, AI as a secondary aid, and explicit instruction in ethical paraphrasing for sustainable academic writing.
Deep & Chen (2025)	Narrative review (SANRA-guided)	20 peer-reviewed studies (2023–2025); higher education students and faculty	ChatGPT; Grammarly; QuillBot; Zotero; Mendeley	AI writing tools enhance writing fluency, grammar, feedback speed, self-directed learning, and cognitive load reduction, but risk over-reliance, weakened critical thinking, plagiarism, loss of originality, and integrity violations. Effective integration depends on pedagogical scaffolding, ethical guidance, transparency, and curriculum redesign. The review emphasizes balanced human–AI collaboration for sustainable academic writing development.
Dergaa, Chamari, Zmijewski & Ben Saad (2023)	Quasi-qualitative literature review	Academic research and publishing context; higher education and scientific writing	ChatGPT (GPT-3), NLP-based AI tools (BERT, RoBERTa, QuillBot, Paperpal)	AI-based writing tools can dramatically improve academic efficiency, including summarization, idea generation, literature scanning, and drafting support. However, they pose significant threats to authenticity, credibility, authorship transparency, plagiarism, and misinformation, including the risk of AI-generated fake science. The study strongly argues that AI cannot replace human researchers, must never be credited as an author, and requires strict transparency, human verification, ethical governance, and accountability. Sustainable academic practice depends on human-AI collaboration, not substitution.
Duah & McGivern (2024)	Qualitative empirical study (semi-structured interviews; inductive thematic analysis)	4 university students and 1 lecturer; UK higher education	ChatGPT; Generative AI tools	Generative AI blurs authorial identity and traditional definitions of academic misconduct, especially in summative assessment. Students are generally more open to AI as a support tool, while educators are more cautious. Both groups highlight policy gaps and ethical uncertainty. The study stresses the urgent need for clear institutional AI usage policies and dialogue between students and faculty to sustain academic norms.

Ely & Rezvani Rad (2024)	Case study / classroom-based empirical study	Undergraduate engineering students; technical writing course (USA)	ChatGPT; Grammarly; Hemingway (generative & non-generative AI)	AI tools enhance grammar, structure, feedback speed, and revision quality, supporting professional writing development. However, AI-generated text often lacks voice, nuance, contextual accuracy, and numerical reliability, and may constrain creativity. The study recommends structured AI literacy modules, clear usage boundaries, and ethical instruction for sustainable integration in writing-intensive courses.
Gallent-Torres, Zapata-González & Ortego-Hernando (2023)	Narrative literature review / conceptual analysis	Higher education institutions globally; students, faculty, and institutions	ChatGPT; Humata.ai; Sudowrite; Generative AI tools	Generative AI offers opportunities such as instant feedback, adaptive learning, accessibility, and resource generation, but raises major ethical challenges related to reliability of information, transparency, data privacy, sustainability, and new forms of plagiarism and authorship impersonation. The authors argue that sustainable AIW integration requires clear ethical guidelines, institutional integrity policies, and pedagogical practices that guide responsible use.
Gustilo, Ong & Lapinid (2024)	Cross-sectional empirical survey grounded in TAM	100 graduate students and faculty across 10 disciplines; Philippines	ChatGPT; Grammarly; Google Translate; QuillBot (Algorithmically-Driven Writing Tools)	Educators perceive high usefulness of AI writing tools for teaching, learning, and assessment, but face barriers including limited access, low AI self-efficacy, and serious concerns about academic integrity, creativity, and output quality. The absence of explicit institutional policies intensifies ethical ambiguity. The study calls for collaborative governance, AI literacy, and policies that privilege human intelligence over automation to ensure sustainable academic practice.
Imran & Almusharraf (2023)	Systematic literature review	30 peer-reviewed studies; global higher education	ChatGPT (GPT-3.5 & GPT-4)	ChatGPT improves writing efficiency, idea generation, language accuracy, and collaboration. However, significant challenges include plagiarism, fabricated citations, assessment integrity, and lack of clear institutional policies. The authors emphasize AI as a supportive aid rather than a replacement for human writing, calling for policy reform and AI literacy training to ensure ethical and sustainable adoption.
Kadwa (2025)	Conceptual / analytical research article (ethical and pedagogical analysis)	Higher education institutions; written assignments; multilingual and second-language contexts	Generative AI tools (e.g., ChatGPT)	AI-based writing tools provide valuable linguistic and structural support, particularly for second-language learners, but raise serious concerns related to academic integrity, transparency, bias, misinformation, equity, authorship, and assessment validity. Traditional written assignments are especially vulnerable. The study recommends process-oriented assessments, reflective portfolios, oral

				defenses, mandatory AI disclosure, and AI literacy education to ensure ethical, fair, and sustainable academic practices.
Koos & Wachsmann (2023)	Conceptual / analytical study (legal education focus)	Higher education, particularly law students and academic examinations	ChatGPT; GPT-4	AI tools support idea generation, structuring, language quality, and productivity, benefiting students with language barriers. However, risks include plagiarism, hallucinated sources, reduced creativity, and erosion of critical thinking. The paper emphasizes ethical awareness, regulation, assessment redesign, and skill development to ensure responsible and sustainable AI-assisted academic writing.
Kotsis (2025)	Qualitative thematic synthesis (review) of 33 peer-reviewed studies (2022–2025)	Higher education academic writing contexts; interdisciplinary	ChatGPT (LLM-based generative AI)	ChatGPT can function as a writing coach, supporting brainstorming, drafting, revision, and multilingual learners. However, risks include algorithmic dependence, fabricated content, plagiarism, unequal access, and erosion of authorship. The study emphasizes AI literacy, transparency, disclosure, equitable institutional policies, and critical pedagogy as prerequisites for responsible and sustainable AI-assisted academic writing.
Kovari (2025)	Opinion / conceptual paper (best-practice framework)	Higher education institutions; educators and students	ChatGPT; generative AI tools	Highlights serious risks of AI-induced plagiarism, hallucinated references, lack of originality, and weakened critical thinking. While ChatGPT enhances accessibility, engagement, and personalized feedback, its misuse threatens academic integrity. The paper proposes best practices such as transparent AI disclosure, adaptive and reflective assessments, oral exams, creative tasks, peer review, and strengthened plagiarism detection. Emphasizes ethical governance, AI literacy, and human oversight as essential for sustainable academic writing.
Mokhine-Martins, Deka & Costa (2024)	Systematic mapping review protocol (PRISMA-guided)	African communities; culturally grounded well-being research	*Not AI-writing focused* (conceptual relevance only)	Although not directly focused on AI-based writing, this protocol contributes a sustainability and ethics lens by integrating Ubuntu philosophy with positive psychology, emphasizing community, equity, inclusion, and cultural relevance. Its principles support human-centred, ethically grounded approaches applicable to AI governance in higher education, particularly in Global South contexts.
Oates & Johnson (2025)	Mixed-methods empirical study (quasi-experimental; rubric-based assessment; focus	10 Master's students in biomedical science programmes; Leeds Beckett University (UK)	ChatGPT	ChatGPT demonstrated strong grammatical accuracy and structural coherence, but did not improve overall essay marks. However, when students critically evaluated AI-generated essays, their critical evaluation scores improved significantly. Students preferred writing their own essays and viewed ChatGPT as a scaffolding and

	group; questionnaire)			evaluative tool rather than a writing substitute. Sustainable AIW use lies in pedagogical designs that promote critical thinking, human oversight, and reflective engagement.
Sarin & Kimkong (2024)	Conceptual / policy analysis	Cambodian higher education context	ChatGPT; Grammarly; QuillBot; Google Translate; LLMs	Highlights urgent need to develop or revise academic integrity policies due to rising risks of plagiarism, cheating, reduced critical thinking, privacy issues, and unreliable AI outputs. Argues that banning AI is unsustainable; instead, institutions should adopt policy-driven, ethical, and pedagogically guided AI use to support long-term academic sustainability, particularly in developing contexts.
Sterling, Ye, Ying & Chen (2025)	Qualitative exploratory study / reflective exercises	International undergraduate and postgraduate students; computing disciplines	ChatGPT (GPT-4); Claude 3.5; Llama 3 (LLMs)	Generative AI helps international students overcome language barriers, blank-page syndrome, and drafting difficulties, improving fluency and efficiency. However, overuse leads to homogenised writing, loss of authorial voice, exaggerated claims, and authenticity concerns. Sustainable AIW use requires careful prompt design, reflective practice, transparency, and explicit instruction on preserving personal voice, positioning AI as a scaffold rather than a substitute.
Turingan (2025)	Sequential explanatory mixed-methods study (survey + interviews)	326 tertiary students (quantitative) + 12 interviewees; higher education	ChatGPT; Grammarly; QuillBot; AI writing tools	Students generally perceive AI writing tools as useful for improving efficiency, organisation, and grammar, supporting perceived usefulness and ease of use. However, major concerns include over-reliance, plagiarism, ethical misuse, and inconsistent institutional AI policies. Significant differences in attitudes were found based on sex, academic year level, and frequency of use. The study emphasizes the need for AI literacy programs, ethical frameworks, and institutional guidelines to ensure sustainable and responsible AI-based academic writing.
Zhao, Cox & Chen (2025)	Empirical quantitative study (survey informed by AI-literacy framework)	124 university students with disabilities; UK higher education	ChatGPT; generative AI chatbots; rewriting tools	Generative AI is widely used by students with disabilities for proofreading, rewriting, summarising, and academic writing support. Benefits include accessibility and learning support, while major concerns include inaccuracy, hallucinations, academic integrity risks, subscription costs, and lack of institutional guidance. Students expressed strong demand for AI literacy training, inclusive policies, and involvement in AI policymaking, highlighting AI's potential for inclusive and sustainable education when governed ethically.

2

4.2.2 Methodological Distribution of Included Studies

Across the final dataset of 27 included studies, systematic and narrative review papers account for approximately 30% of the literature (n = 8). These review studies were intentionally included to capture synthesised evidence, theoretical discussions, and ethical debates in a rapidly developing field. Given the early stage of research on AI-based writing tools in higher education, review papers play an important role in identifying overall trends, conceptual gaps, and emerging sustainability concerns. Empirical studies were mainly used to support conclusions related to learning outcomes and practical use.

Among the empirical studies, cross-sectional survey designs represent 22% of the dataset (n = 6), indicating a strong focus on perceptions, attitudes, and adoption behaviours of AI-based writing tools. Qualitative studies account for 19% (n = 5) and provide deeper insights into user experiences, ethical concerns, and pedagogical implications. Mixed-methods studies also make up 19% (n = 5), reflecting increasing efforts to combine quantitative and qualitative evidence. In contrast, experimental or quasi-experimental studies are relatively limited, representing only 13% of the included studies (n = 3).

Overall, this distribution suggests that research on AI-based writing tools in higher education is still largely exploratory and descriptive in nature. There remains a clear lack of rigorous intervention-based studies that examine causal effects and long-term learning outcomes. As a result, claims about pedagogical effectiveness are interpreted with caution when they are based mainly on cross-sectional or self-report data. Figure 3 illustrates the distribution of study designs across the included literature.

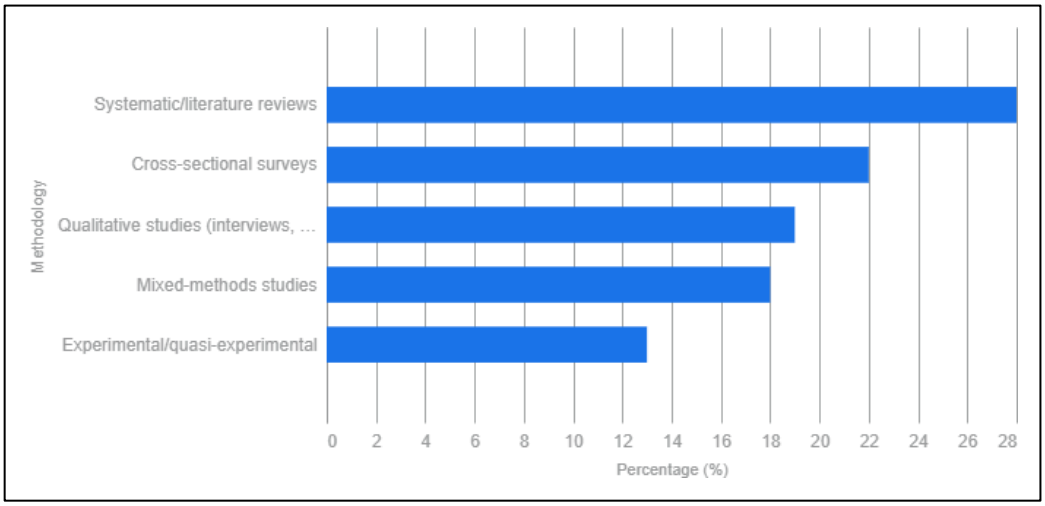

Figure 3: Methodology Distribution

4.2.3. Geographical Distribution

The included studies show an uneven geographical distribution across global regions. Around 32% of the studies (n = 9) were conducted in North America, followed by Europe with 28% (n = 8) and the Asia–Pacific region with 24% (n = 6). Research from the Middle East accounts for 11% of the studies (n = 3), while Africa is represented by only one study (3%). In addition, a small number of studies were conducted across multiple countries or did not specify a single geographical context, most often in large-scale review or conceptual papers.Overall, this pattern shows that research on AI-based writing tools in higher education is largely concentrated in the Global North. In contrast, evidence from under-resourced and Global South regions remains limited. This imbalance is important, as several review and multi-country studies included in this review highlight the need for broader geographical coverage to better reflect differences in educational systems, infrastructure, language contexts, and equity conditions. The findings therefore point to a clear geographical research gap and underline the need for more empirical studies conducted in diverse cultural, educational, and socio-economic settings to support more

inclusive and sustainable academic practices. Percentages are rounded to the nearest whole number (see Figure 4).

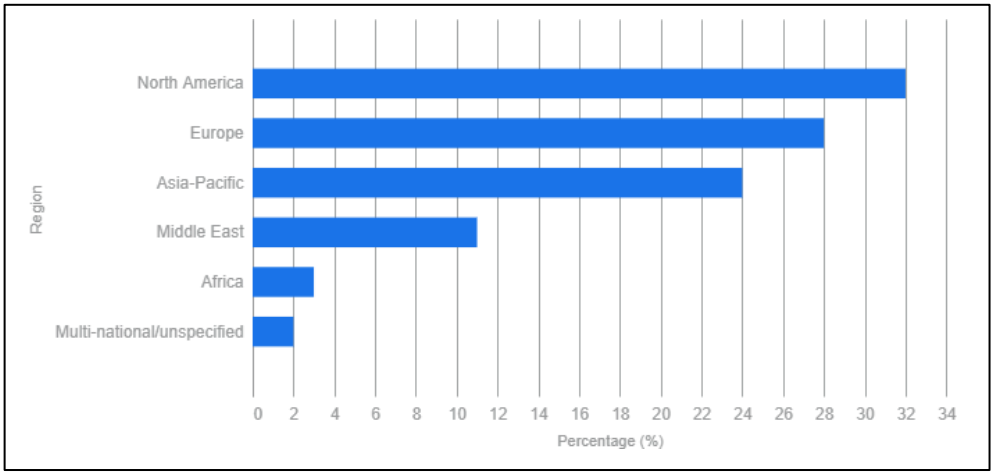

Figure 4, Geographical Distribution

4.2.4 AI-Based Writing AI Tools Examined

Regarding the technologies examined, ChatGPT and related large language models (e.g., GPT-3.5 and GPT-4) dominate the literature, appearing in approximately 67% (n = 18) of the included studies. In addition, around 30% (n = 8) of studies refer more broadly to multiple or unspecified generative AI tools, reflecting a growing focus on general AI-assisted writing practices rather than single platforms.

By comparison, non-generative or task-specific writing tools are discussed less frequently, including Grammarly (18%, n = 5), QuillBot (12%, n = 3), and other specialist academic writing assistants such as Writefull, Scite, and Jasper (8%, n = 2). As several studies examined more than one tool, these categories are not mutually exclusive (see Figure 5).

Overall, this distribution shows that while ChatGPT has received considerable research focus, other AI-based writing tools are relatively less researched, which presents an opportunity for more diversified research in future studies.See figure 5.

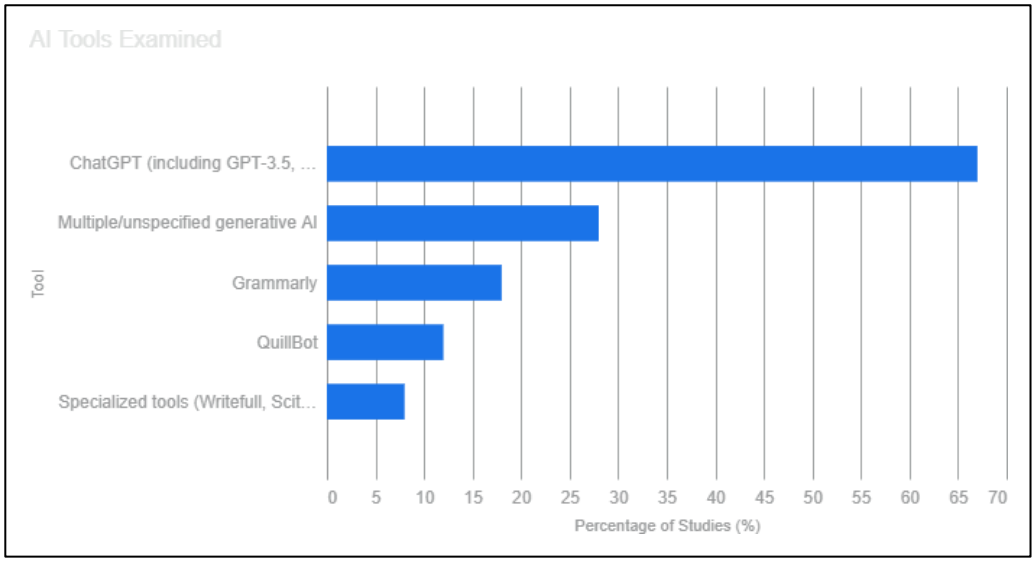

Figure 5. General AI Writing Tools

4.3 Methodological Quality Overview

The methodological quality of the included studies was assessed using the Mixed Methods Appraisal Tool (MMAT, 2018). Overall, the quality of the evidence ranged from moderate to high, with a smaller number of studies assessed as low quality due to limitations in sampling, study design, or analytical transparency.

Most studies met the MMAT criteria related to clarity of research questions and alignment between research aims, methods, and outcomes. Studies rated as high quality typically provided clear methodological descriptions, appropriate data collection procedures, and well-justified analytical approaches. Moderate-quality studies often relied on self-report data or cross-sectional designs, which limited causal inference but still offered valuable descriptive and contextual insights. A small number of low-quality studies showed weaknesses such as limited sample justification or insufficient detail in data analysis.

Methodological quality was not used as an exclusion criterion. Instead, MMAT assessments were used to inform the level of confidence placed in different findings during synthesis. Evidence from higher-quality studies was given greater weight when drawing conclusions, particularly for claims related to pedagogical effectiveness and learning outcomes, while findings from lower-quality studies were interpreted more cautiously.

All the scores of the included studies in terms of the detailed evaluation of MMAT (2018) quality are provided in Supplementary File C, which enables readers to assess the methodological quality of the study on their own. During the main analysis, the quality of the studies was taken into account when interpreting the results whereby more consideration was given to patterns that were supported by higher-quality studies. Poor quality studies were not discarded but were viewed with caution especially where the results were not backed by the high-quality studies. Quality appraisal using MMAT was conducted independently by two reviewers, with discrepancies resolved through discussion and consensus.

4.4. Thematic Synthesis of Findings

Across the 27 included studies, pedagogical benefits were reported in 18 studies, most commonly in relation to writing efficiency, feedback, and language support. Ethical and integrity challenges were identified in 21 studies, making this the most consistently reported theme, particularly in relation to plagiarism, over-reliance, and authorship ambiguity. Institutional governance and sustainability concerns were raised in 16 studies, with emphasis on policy development, AI literacy, and assessment redesign. Inclusivity and accessibility were explicitly addressed in 6 studies, primarily focusing on support for non-native English speakers and students with disabilities.

Together, these patterns indicate that current knowledge about AIW tools in higher education is shaped largely by early-adopting institutions in the Global North, relies heavily on self-report and cross-sectional designs, and focuses predominantly on generative systems such as ChatGPT.

Based on the thematic synthesis, four higher-order themes were identified: pedagogical benefits, ethical and integrity challenges, inclusivity and accessibility, and institutional governance and sustainability (Figure 6). Details of the coding framework and theme extraction process are provided in Supplementary File D, supporting transparency of the synthesis. Initial open coding generated descriptive sub-themes (e.g., writing efficiency, academic integrity, accessibility, governance), which were subsequently grouped into higher-order themes through iterative comparison and refinement. This hierarchical structuring ensured conceptual coherence between the extracted data and the final thematic framework.

From a theoretical perspective, these patterns can be interpreted through sociocultural and self-regulated learning frameworks. AIW tools function as scaffolding mechanisms when used strategically, enabling learners to refine their work within their developmental capacity. However, when these tools replace

rather than support cognitive processes, they may undermine learning. Across themes, benefits were most evident where tools operated as scaffolds with gradual fading and where learners demonstrated strategic self-regulation, whereas risks were more pronounced in contexts characterized by automation without fading and dependent use.

Conflicting findings across studies were retained and interpreted in relation to methodological quality, study context, and sample characteristics. This approach enabled the identification of both convergent and divergent patterns within the evidence base.

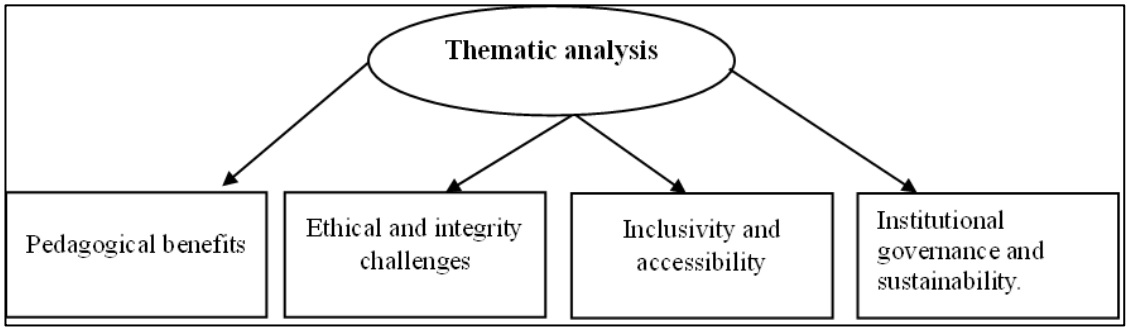

Figure 6. Higher-Order Themes Derived from Thematic Synthesis of the Included Studies

4.4.1. Theme 1: Pedagogical Benefits and Learning Outcomes

The literature documents several pedagogical advantages of AIW tools, with important caveats on instructional design and implementation.

1. Immediate Feedback and Writing Improvement

Perhaps the most highly cited advantage of using AIW tools is their capability to provide instant feedback and correction on grammar, coherence, structure, and style. Many studies show that students, especially multilingual learners, use AIW tools to refine writing drafts and improve accuracy and coherence. For example, Angelos et al. (2024) found that 89% of the Saudi EFL learners in their study reported improved academic writing skills after integrating the use of ChatGPT into the process of writing revision. They also found enhanced confidence among the learners in expressing complex ideas.

2. Generation of Ideas and Creativity Enhancement

AIW tools have been found to be useful in the brainstorming stage. For example, AlAfnan's (2023) study found that ChatGPT can generate multiple examples, alternative phrasings, and useful conceptual frameworks to help students and instructors as a starting point for their written work. Konstantinos (2025) synthesized 33 studies and found that ChatGPT was useful as a writing coach to help students overcome writing anxiety and explore different approaches for academic writing. This scaffolding, in the form of content ideas, appears especially valuable for novice writers and those unfamiliar with academic writing in their disciplines. However, researchers caution against overreliance on AI-generated ideas as they may constrain rather than expand creative thinking if students adopt suggestions uncritically rather than using them as spring boards for original thought (Ely & Rad, 2024).

3. Scaffolding and Personalized Learning Support

Several studies position AIW tools as cognitive scaffolds that can be gradually withdrawn as learners develop competence. King et al.'s (2024) mapping review across 22 countries found consistent evidence that generative AI provides on-demand tutoring, offering explanations of writing conventions, modeling revision strategies, and responding to student queries, functions traditionally limited by instructor availability and class size. Kotari (2025, p.7) suggests that "another pedagogical method that AI has

changed in collaborative environments involves project-based learning. It requires learners to work in groups and solve complex real-world problems with continuous feedback, recommendations, and resources provided by these AI systems." This illustrates the learning support provided by AI systems to student learning, including for writing tasks.

4. Efficiency and Workflow

Past studies consistently found that AIW tools helped both lecturers and students save time when carrying out writing tasks such as drafting outlines, organizing citations, and summarizing materials. This helps free up cognitive resources for higher-order thinking skills such as analyzing, evaluating, and synthesizing past studies, making inferences, and drawing conclusions. For example, Gustilo et al.'s (2024) survey of 100 Filipino lecturers and graduate students found that perceived usefulness was the strongest predictor of adoption intention, with efficiency benefits outweighing concerns.

However, critics have argued that efficiency is achieved at the expense of learning. Critics within the reviewed literature argue that the productive challenges inherent in writing, which are part of the process for developing writing skills such as dealing with ambiguity, revising repeatedly, and developing voice, may be shortchanged when AIW tools make writing "too easy" (Duah & McGivern, 2024). This tension underscores the need for pedagogical intervention. Efficiency is pedagogically valuable when it eliminates busy work, but it is problematic when it bypasses learning processes.

5. Development of Critical Evaluation Skills

Recent studies suggest that integrating AIW tools into pedagogical practices could enhance critical thinking rather than diminish it. A classroom-based study by Oates and Johnson (2025) showed that teaching students to evaluate ChatGPT outputs for accuracy, bias, and appropriateness would lead to strong evaluation and argumentation skills. This aligns with calls for AI literacy to be taught to students as part of the aim to promote critical and ethical use of AI tools in tertiary education.

4.4.2. Theme 2: Ethical and Academic Integrity Challenges

Though pedagogical benefits emerge from the positive findings shown by past literature, there is an ongoing debate on ethical and integrity concerns in the use of AIW tools for academic writing.

1. Plagiarism and Paraphrasing Concerns

Just by providing prompts and questions, AIW tools such as ChatGPT can generate essays that are coherent and sophisticated. Students can pitch the generated essays to a level they require and further prompt the AIW tool to refine the writing. The capacity of AIW tools to produce texts for a multitude of topics and purposes has resulted in students taking this shortcut to academic writing.

2. Authorship and Attribution Ambiguities

The issue of authorship is also a primary concern. To what extent can AIW tools be used for a student to still be able to claim authorship of a piece of writing? When ChatGPT generates an outline, refines the arguments, and edits the writing, who is the author? These are concerns for lecturers, students, and higher education governing bodies. There appears to be a contradiction in the understanding of the use of AI among lecturers and students. Duah and McGivern's (2024) qualitative interviews with UK students and lecturers indicate that students described AI as a tool, much like Google, but lecturers tend to view the use of AI as academic dishonesty, as it undermines authenticity and student ownership of writing.

3. AI Detection Challenges

The literature documents growing scepticism about the use of AI-detection tools. These tools are often seen as producing false positive or false negative rates. The different tools also often show discrepancies in the detection of AI-generated text, particularly for non-native English speakers whose writing may be flagged as AI-generated due to stylistic patterns (Kővári, 2025). AlAfnan et al. (2023) found that even straightforward paraphrasing requests to ChatGPT rendered text undetectable by current tools. Thus, the reliability and trustworthiness of these AI-detection tools are highly doubted. However, Kővári's (2025) comprehensive review notes that detection tools may still play transitional roles in institutional responses, particularly when combined with process-based verification methods.

4. Misinformation, and Fabrication

Generative AI has been known to generate fabricated information or what is termed as "hallucinations". The generated content sounds plausible but actually contains fabricated data, false citations, made-up references, and false claims. A number of studies have shown that students unknowingly incorporate these fabricated references and content into their academic writing (AlSofi, 2024; Kotsis, 2025). When students do not fact-check their AI-generated content, there is the likelihood that they will accept this confident-sounding output generated by AIW tools. This problem underscores the need to teach AI literacy in tandem with teaching writing skills. Table 2 summarizes the challenges in academic integrity and proposed strategies to address these challenges.

Table 2. Academic Integrity Challenges and Proposed Mitigation Strategies

Challenge	**Evidence**	**Proposed Mitigation Strategies**	**Limitations**
Paraphrasing plagiarism	ChatGPT-produced text can bypass similarity checkers (AlAfnan et al., 2023)	• Redesign assessments (contextual, authentic writing tasks) • Require process portfolios and drafts • Enforce disclosure of AI use	Detection tools unreliable; requires major pedagogical shifts
Authorship ambiguity	Students more accepting of AI authorship; staff uncertain without policy (Duah & McGivern, 2024)	• Establish institutional authorship policies • Develop transparent AI acknowledgment norms • Integrate academic integrity and AI ethics training	Persistent philosophical disagreement; policies may not align across disciplines
Detection failures	False positives and negatives common in AI detection systems (Kővári, 2025)	• Reduce reliance on AI detection tools alone • Emphasize iterative writing and process assessment • Use oral defenses or viva checks where relevant	Process-based approaches require staff training and time; institutional capacity varies
Data privacy	Risks of student data being exposed to third-party AI platforms (Costa et al., 2024)	• Use institutionally licensed AI tools • Implement robust data governance and privacy protocols • Train students to critically evaluate platform privacy policies	Institutional solutions costly; limited control over commercial AI data practices
Bias and misinformation	AI may generate biased content and fabricated citations (AlSofi, 2024; Kotsis, 2025)	• Teach critical AI literacy and verification skills • Require citation checks and fact validation steps • Encourage diverse, well-designed prompts	Lecturer and student AI skills vary; risk evolves as models update

Note: Evidence derived from included studies; table summarises reported strategies rather than evaluated effectiveness.

4.4.3. Theme 3: Inclusivity and Accessibility

AIW tools present complex implications for educational equity. Though they offer support for multilingual learners, they raise issues of lack of access for marginalized learners, which further exacerbates concerns of inequality in education.

1. Support for Multilingual and Non-Native English Writers

Support for non-native and multilingual speakers has been identified as a strength afforded by AIW tools. This support includes providing corrections in grammar use, using appropriate idioms, lexical items, and academic register (Costa et al., 2024). AlSofi's (2024) study of Saudi EFL undergraduates revealed that ChatGPT use positively influenced writing confidence and willingness to attempt complex sentence structures. This would suggest that AI support may reduce anxiety and encourage risk-taking among language learners. Kotsis's (2025) synthesis similarly found AI as especially beneficial for multilingual and novice writers during brainstorming and revision phases.

2. Assistance for Students with Disabilities

AIW tools have also been found to facilitate students with disabilities in their writing journey. A survey by Zhao et al. (2024) found that 124 students with ADHD, dyslexia, dyspraxia, and autism spectrum conditions in UK higher education actively used generative chatbots and rewriting apps to overcome writing-related barriers, organize thoughts, maintain focus and produce coherent drafts. This would point to the assistance that AIW tools provide for writing tasks among students with certain disabilities.

3. Equity Considerations and Access Barriers

Though AIW tools appear to provide access to writing support for students in higher education, the review of past studies show some threats in equity. Gustilo et al.'s (2024) Philippine study found that access barriers such as limited internet connectivity, lack of institutional subscriptions to AIW tools, and low digital literacy could adversely affect students from low socioeconomic backgrounds. In the same vein, Zhao et al. (2024) reported that premium AIW tool subscriptions (e.g., ChatGPT Plus, Grammarly Premium) created a dividing system where affluent students accessed more powerful features while others made do with limited free versions. Costa et al. (2024) found that these access disparities are consistent with existing educational inequalities, where under-resourced institutions that lack infrastructure can only provide limited access to AIW tools, while well-funded institutions can offer premium licenses and comprehensive training. This pattern risks widening rather than narrowing global educational divides.

4. Representation and Cultural Responsiveness

An emerging concern in the literature revolves around AIW tools' capacity for cultural responsiveness. Given that generative AI systems appear to be more in tune with the English language and western academic texts, it may default into western cultural assumptions and norms, thus marginalizing non-western knowledge systems, writing traditions, and writing conventions (Costa et al., 2024). Thus, students from non-Western backgrounds might feel disconnected from the AIW contexts. When students rely on AIW tools to generate examples, discussions, and arguments, they may receive generated content with a dominance of Western culture and perspectives, thus sidelining non-Western viewpoints. This is especially salient in fields such as humanities, arts and social sciences, where diverse and contradicting perspectives are common.

4.4.5 Theme 4: Institutional Governance and Sustainability

The sustainability of AIW tools also depends on lecturer preparedness, policy development and enforcement, and institutional capacity.

1. Policy Development and Institutional Responses

The reviewed studies highlight concerns among lecturers and students about the lack of clear policies and procedures for the use of AIW tools. Duah and McGivern's (2024, p. 9) study in the UK found that "students may take a more laissez-faire approach to GenAI, viewing it as a complimentary tool and part of the writing process. Academics may view it in a more negative light and a tool which is perhaps more likely to jeopardise authorial integrity." Thus, the absence of clear policies on the use of generative AI systems results in ethical ambiguity and uncertainties among both lecturers and students. Students were unsure of the permissible use of AI, while lecturers showed inconsistent practices in allowing AI use among students. This lack of policy appears to be widespread among higher education institutions. In a similar vein, Costa et al.'s (2024) review found that fewer than 30% of institutions studied had published comprehensive AI use guidelines by mid-2024.

In general, current guidelines in the use of AIW tools require students to disclose how they used generative AI in their writing process (Kővári, 2025). However, collectively, there appears to be a lack of understanding of disclosure requirements and practices. A number of studies adopt the use of the OTHA framework, which encompasses Openness, Transparency, Honesty, and Accountability. This framework emphasizes transparent and explicit acknowledgement of AI-generated content and its use in students' written work. However, how this framework translates into actual practices need to be further investigated and made explicit for lecturers and students.

2. Lecturer Training and Digital Literacy Needs

Lecturer preparedness in incorporating AIW tools and managing students' use of these tools also poses a challenge for their optimal and ethical use. Gustilo et al.'s (2024) survey found that even though Filipino lecturers recognized the usefulness of AIW tools, they reported low self-efficacy in using the tools within the teaching and learning context. They also reported limited understanding of the AIW tools' capabilities and limitations. This study raises several issues, including misconceptions about how AIW tools function and ill-informed policy development. Thus, there is a call for professional development in AI-literacy and AI policy development. Literature suggests that comprehensive lecturer professional development should be carried out and it should include hands-on experience with AIW tools, pedagogical strategies for integrating AIW tools, ethical guidelines for AI use, and AI literacy (e.g., Kővári, 2025; Turingan, 2025).

3. Assessment Redesign for Sustainable Integrity

Rethinking and reformulating assessment practices are key in ensuring ethical and genuine authorship contributions in writing. Rather than prohibiting the use of AIW tools, a more productive approach would be to implement evidence-based strategies (Turingan, 2025). This would include the use of process-based writing, where the writing and revising of the drafts are monitored. Another strategy would be using authentic and contextualized tasks as assignments requiring personal experience, local knowledge, or specific course content are harder to outsource to generative AI tools (AlAfnan et al., 2023). Adding an oral component to the writing task could also ensure academic writing honesty. Pairing written work with oral defenses, presentations, or in-class discussions allows for verification of understanding and reduces the chances for AI misuse (Kővári, 2025). Multimodal and collaborative projects could also help maintain integrity in the written work. Assignments incorporating audio, video, visual, or hands-on components or requiring genuine collaboration are more difficult to complete solely through AI tools (Kővári, 2025). In addition, group projects with peer evaluation add social accountability that discourages free riding through AI. Metacognitive reflection could also be used as it

would require students to reflect on their writing process, including AI tool use. This could foster transparency and critical thinking about technology's role in learning (Turingan, 2025).

4. Infrastructure and Resource Considerations

Sustainable AI integration requires institutional investment in infrastructure such as licenses for AI tools, technical support systems, learning management system integration, and ongoing evaluation mechanisms (Gustilo et al., 2024). The literature suggests that ad hoc, individual adoption, creates inequities and confusion. Thus, there is a need for higher institutions to play active roles in ensuring adequate investment to support the use of AIW tools that would benefit learners. Another aspect that needs attention is the environmental cost we pay as a result of the extensive use of AI systems. Large language models require substantial computational resources and energy, raising questions about the carbon footprint of widespread AIW tool use (Richie, 2025). Only a handful of studies mention environmental sustainability, representing a significant gap given higher education's commitment to climate action and SDG alignment.

5. Alignment with SDG 4: Quality Education Goals

While many studies implicitly connect AIW tools to quality education goals, explicit alignment with UN Sustainable Development Goal 4 remains limited. The reviewed literature suggests several potential connections to SDG 4: Quality Education, through encouraging AI literacy as a skill for employment, as well as promoting equity and inclusivity in education. However, currently these connections remain largely aspirational and vague. The gap between AIW tools' potential to advance SDG 4 and actual implementation that realizes this potential represents a critical area for future research and policy development.

Table 3 summarizes the institutional governance dimensions and implementation challenges highlighted by the reviewed studies. These are the challenges related to SDG 4 and the use of AIW tools.

Table 3. Institutional Governance Dimensions and Implementation Challenges

Governance Dimension	Current State	Challenges	Recommendations
Policy Development	Many institutions lack clear AI use guidelines (Duah & McGivern, 2024)	Rapid AI evolution; disciplinary norms differ; philosophical debates on authorship and integrity	Adopt OTHA governance framework; involve students & lecturers in policy design; review and update policies iteratively.
Lecturer Training	Limited lecturer preparedness; low self-efficacy reported (Gustilo et al., 2024)	Resource and time constraints; uneven motivation; lack of AI expertise	Provide structured professional development; offer discipline-specific case examples; build continuous support systems and communities of practice
Assessment Redesign	Emerging practices in early adoption stages (Kővári, 2025)	Increased workload; academic resistance; limited evidence of longterm outcomes	Prioritize authentic, process-based, and multimodal assessments; share evidence-based models; support gradual implementation
Equitable Access	Unequal AI access across socioeconomic and geographic groups (Costa et al., 2024)	Financial barriers to licenses; tech infrastructure gaps; digital literacy inequality	Institutional AI subscriptions; universal design approaches; targeted digital literacy and support programs.
Student AI Literacy	Early stage; few structured AI literacy initiatives (Turingan, 2025)	Crowded curricula; uneven lecturer preparedness; unclear competency standards	Embed AI literacy across courses; model ethical & critical AI use; evaluate AI skill development

Evaluation and Research	Limited longitudinal and cross-context evidence	Rapid change outpaces research; methodological complexity; resource demands.	Conduct longitudinal and comparative studies; enable cross-institutional research; use mixed methods and iterative evaluation.

Note: Evidence derived from included studies; table summarises reported strategies rather than evaluated effectiveness.

To avoid equating evidence of differing methodological strength, thematic interpretations were examined in relation to study design and quality appraisal. Pedagogical benefit claims were supported by a mix of experimental and quasi-experimental studies (n =5) and self-report survey studies (n = 8), indicating stronger evidence for feedback and language support effects than for long-term learning outcomes. In contrast, many ethical and integrity concerns were derived from conceptual analyses, reviews, and cross-sectional surveys, highlighting perceived rather than experimentally verified risks. Inclusivity-related findings were primarily supported by targeted empirical studies involving students with disabilities, whereas institutional governance claims were largely grounded in policy analyses and synthesis studies. MMAT ratings were used to inform the weighting of evidence, with findings from higher-quality studies receiving greater interpretive emphasis.

5. Discussion

5.1 Overview of Key Findings

This review synthesised evidence from 27 studies on the use of AI-based writing (AIW) tools in higher education. Overall, the findings show a clear pattern: AIW tools offer practical benefits, but they also introduce important risks and challenges.

As shown in Table 4, strong evidence supports the role of AIW tools in improving writing efficiency and feedback, based on multiple high-quality studies.

Table 4. Summary of the Major Analytical Claims of the review, the number and type of supporting studies, and the relative strength of evidence based on MMAT appraisal

Major Claim	Supporting Studies (n)	Dominant Study Types	Evidence Strength
AIW tools improve writing efficiency and feedback	12	Experimental; mixed methods; surveys	**High**
AIW tools enhance grammar and language support	10	Experimental; classroom studies	**High**
AIW tools increase plagiarism and integrity risks	15	Surveys; reviews; conceptual	**Medium**
Over-reliance may erode critical thinking	11	Surveys; conceptual	**Medium**
AIW improves accessibility for disabled students	6	Targeted empirical studies	**Moderate–High**
Sustainable AIW use requires institutional governance	16	Reviews; policy analyses	**Medium**

The most consistently reported benefits include improved writing efficiency, faster feedback, and better language support, especially for non-native English speakers (e.g., Angelos et al., 2024; AlAfnan et al., 2023; Gustilo et al., 2024). At the same time, concerns about academic integrity, over-reliance, and reduced critical thinking appear across a large proportion of studies (e.g., Duah & McGivern, 2024; Kővári, 2025). Issues related to governance, policy, and sustainability are also widely discussed, while inclusivity and accessibility receive comparatively less attention.

Importantly, the strength of evidence varies across themes. Some findings are supported by multiple high-quality studies, while others rely on more limited or mixed evidence. This suggests that conclusions should be interpreted with appropriate caution.

Taken together, the results show that AIW tools are not inherently beneficial or harmful. Their impact depends on how they are used, the context in which they are applied, and the level of support provided to students.

5.2 Geographical and Disciplinary Variations

The review also reveals clear geographical differences in research focus and regulatory concerns. Studies conducted in North America and Europe tend to foreground academic integrity, AI detection, and assessment security, whereas research from Asia–Pacific and Middle Eastern contexts more often highlights pedagogical and equity-related benefits, particularly for language support and access (AlSofi, 2024; Costa et al., 2024; Zhao et al., 2024).

These differences highlight the limitations of uniform governance approaches. In multilingual and under-resourced contexts, AI-assisted writing may function as a legitimate accessibility tool rather than primarily as a threat to academic integrity (Costa et al., 2024). By contrast, in highly resourced and monolingual settings, concerns about authorship attribution and assessment validity appear more pronounced (Kővári, 2025). Effective governance therefore requires context-sensitive policy frameworks rather than universal prohibition or uncritical adoption.

Disciplinary variation remains insufficiently examined. Most studies treat academic writing as a general skill, despite clear differences across STEM, humanities, and professional disciplines. This represents a significant gap, as the pedagogical implications of AIW use are likely to vary across disciplinary epistemologies and assessment traditions (Oates & Johnson, 2025).

5.3 Theoretical Interpretation: Why and When AIW Tools Work

The findings can be better understood through the perspectives of Vygotsky's Zone of Proximal Development (ZPD) and Zimmerman's self-regulated learning theory.

From a sociocultural perspective, AIW tools can act as mediating supports that assist learning within the learner's ZPD. Features such as automated feedback and writing suggestions can provide structured guidance that helps students refine their work beyond their current independent ability. However, this support is effective only when learners actively engage with it. When students accept AI-generated outputs without reflection, the tool bypasses the developmental processes central to learning.

Self-regulated learning theory further explains this variation. Students who use AI tools strategically, by evaluating, revising, and integrating suggestions, tend to benefit more (Zimmerman, 2002). In contrast, students who rely on AI passively are more likely to develop dependency and show reduced critical engagement.

This explains why the same tools produce different outcomes across contexts. Positive outcomes occur when AI supports cognitive processes, whereas negative outcomes emerge when it replaces them.

5.4 Balancing Efficiency and Learning

One of the most frequently reported benefits of AIW tools is efficiency. These tools can reduce the time required for drafting, editing, and language correction (Gustilo et al., 2024; Ely & Rad, 2024). This is particularly helpful for novice writers and multilingual learners.

However, efficiency should not be treated as an unquestioned advantage. Writing is not only a product but also a cognitive process that involves reflection, organization, and revision. These processes require effort and are essential for learning.

When AI tools reduce effort too much, they may also reduce learning. Students may produce better texts but engage less deeply with the content. This creates a tension between speed and understanding.

At the same time, efficiency can support learning when it reduces lower-level cognitive load. For example, grammar correction tools can free time for higher-order thinking. Therefore, the impact of efficiency depends on whether it supports or replaces meaningful engagement.

5.5 Differentiated Effects by AIW Tool Type

A key contribution of this review is the distinction between different types of AIW tools. These tools differ in function, and therefore in their pedagogical value and risks.

Grammar and style tools, such as Grammarly, operate on student-produced text and provide corrective feedback. These tools support revision and align with process-based writing pedagogy, particularly for multilingual learners (e.g., Zawacki-Richter et al., 2019).

Paraphrasing tools, such as QuillBot, can assist with language expression but may lead to superficial rewriting if used without understanding. This raises concerns about reduced comprehension and weak synthesis skills.

Generative AI systems, such as ChatGPT, represent a more fundamental shift. These tools can produce complete text outputs, potentially bypassing key stages of thinking and writing. While they may support brainstorming, they also raise concerns about authorship and academic integrity (e.g., Kasneci et al., 2023).

Overall, the evidence suggests that AIW tools should not be treated as a single category. Their educational impact depends on their function, and different tools require different pedagogical and policy approaches.

5.6 Equity, Power, and Structural Implications

AIW tools also raise important questions about equity and power. While they are often presented as tools that democratize access, their benefits are not equally distributed.

First, many tools are aligned with standardized academic English, which may disadvantage students with diverse linguistic backgrounds (Zawacki-Richter et al., 2019). This can reinforce existing norms rather than support linguistic diversity.

Second, access to advanced AI tools is often linked to cost. Paid features and subscription models may create inequalities between students with different financial resources (Zhao et al., 2024; Costa et al., 2024).

Third, AIW tools operate within commercial systems where user data may be collected and used. This raises concerns about privacy, ownership, and the commodification of student work (Kasneci et al., 2023).

These issues suggest that AIW integration is not only a pedagogical matter but also a social and institutional one. Addressing these concerns requires critical and inclusive approaches.

5.7 Sustainability Considerations

AIW tools are likely to remain part of higher education in the long term, which raises important sustainability questions.

From a social perspective, sustainability involves equitable access, responsible use, and long-term learning outcomes. Institutions need to ensure that all students can benefit from these tools.

Environmental sustainability is also an emerging concern. Large AI systems require substantial computational resources, contributing to energy use and carbon emissions (Ritchie, 2025). This issue has received limited attention in educational research but is increasingly relevant.

Institutions should therefore consider sustainability when adopting AI tools, including responsible procurement and usage practices.

5.8 Policy Implications

The findings of this review highlight the need for clear and actionable policies.

One of the key implications is that institutions should differentiate between tool types and develop appropriate guidelines for their use. Though grammar tools can be broadly permitted as they generally deal with language accuracy, generative AI tools require more careful regulation (Kasneci et al., 2023).

Ensuring transparency in AI use is also essential. Students should disclose how AI tools are used in their work. This would ensure academic integrity and promote ethical use of generative AI.

Assessment practices would also need to evolve. Process-based assessments, such as writing drafts and conducting reflections, can better capture student learning and help circumvent unethical and overuse of AI tools.

Furthermore, training is critical for both students and educators. Suitable training sessions should be conducted to provide guidance on responsible AI use.

Finally, given the rapid evolution of AI tools, policies should be regularly reviewed to keep pace with technological change.

5.9 Limitations

This review has several limitations. It includes only peer-reviewed studies, which may exclude relevant insights from grey literature. This may also introduce publication bias, as positive findings are more likely to be published.

In addition, the number of included studies is relatively small, and many rely on self-reported data. This might limit the strength of the conclusions.

5.10 Future Research Directions

Future research should examine long-term effects of AIW tools on writing development. More experimental and longitudinal studies are needed.

Research should also explore differences across student populations, disciplines, and contexts. In addition, more work is needed on institutional implementation and policy effectiveness.

6. Conclusion

AI-based writing (AIW) tools are rapidly reshaping academic writing practices in higher education. This review shows that these tools offer clear benefits, including improved writing efficiency, immediate feedback, and enhanced language support, particularly for multilingual learners. At the same time, they introduce important challenges related to academic integrity, over-reliance, and the potential weakening of critical thinking and writing development.

The findings highlight that the impact of AIW tools is not determined by the technology itself, but by how it is used within pedagogical and institutional contexts. When integrated thoughtfully, AIW tools can function as supportive learning aids that enhance student engagement and development. However, when used without guidance, they may replace essential cognitive processes and undermine meaningful learning.

This review also demonstrates that AIW tools should not be treated as a single category. Different tools serve different functions and therefore require differentiated pedagogical approaches and policy responses. In addition, issues of equity, access, and sustainability must be considered to ensure that the benefits of AI are distributed fairly and responsibly.

Overall, the challenge for higher education is not whether to adopt AIW tools, but how to integrate them in ways that support learning, uphold academic integrity, and promote inclusive and sustainable educational practices. Thoughtful design, clear policies, and ongoing evaluation will be essential to ensure that AI serves as a tool for learning rather than a substitute for it.

Funding

This research is funded by the Research Incentive Grant for Teaching and Learning (GIPP), Universiti Putra Malaysia, grant no. 9323840.

Acknowledgments

The authors would like to thank the anonymous reviewers of this article. The authors would also like to acknowledge the support of the university-driven research program on lifelong learning and reskilling.

Declaration

The authors would like to declare that AI tools were used solely for language editing and clarity enhancement. No AI tools were used for data extraction, thematic coding, quality appraisal, synthesis, or interpretation of findings. All analytical decisions were made by the authors.

Data Availability Statement

The data extraction spreadsheet and the coding framework used for the thematic synthesis are available as supplementary materials. To support transparency and replicability, anonymised versions of these materials have also been deposited in an open-access repository, which can be accessed at: **[DOI: 10.17605/OSF.IO/F8P6C]**.

References

AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology, 3*(2). https://doi.org/10.37965/jait.2023.0184

Almusharraf, M. I. N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology.* https://doi.org/10.30935/cedtech/13605

Angelos, C., Antonia, S., Athanasia, K., Maria, K., Anastasia, M., Marina, M., & Sofia, M. (2024). Impact of artificial intelligence and virtual reality on educational inclusion: A systematic review of technologies supporting students with disabilities. *Neveléstudomány.* https://doi.org/10.3390/educsci14111223

Bui, T. T. U., & Tong, T. V. A. (2025). The impact of AI writing tools on academic integrity: Unveiling English-majored students' perceptions and practical solutions. *AsiaCALL Online Journal, 16*(1), 83–110.

Carhuaz Valdez, D. D., & Arcata-Maquera, E. J. (2024). *University education in the era of generative artificial intelligence: Is academic integrity at risk? Ingeniería Investiga*, *6*. https://doi.org/10.47796/ing.v6i00.967

Charmaine, B. (2023). To use or not to use ChatGPT and assistive artificial intelligence tools in higher education institutions? The modern-day conundrum – students' and faculty's perspectives. *Equity in Education & Society.* https://doi.org/10.1177/27526461231215083

Costa, K., Ntsobi, P. M., & Mofolo, L. (2024). Challenges, benefits and recommendations for using generative artificial intelligence in academic writing: A case of ChatGPT. *Medicon Engineering Themes, 7*(4).

Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In D. A. Buchanan & A. Bryman (Eds.), *The SAGE handbook of organizational research methods*. Sage Publications.

Duah, J. E., McGivern, P. J., & Technology, L. (2024). How generative artificial intelligence has blurred notions of authorial identity and academic norms in higher education, necessitating clear university usage policies. *International Journal of Information and Learning Technology, 41*(2), 180–193.

Ely, S. E., & Rad, M. (2024). Examining the opportunities and challenges of using artificial intelligence for engineering technical writing courses. https://doi.org/10.18260/1-2--47381

Flores, R. E. A., & Trujillo, S. M. (2025). Investigación científica e inteligencia artificial en estudiantes de posgrado: Un análisis cualitativo [Scientific research and artificial intelligence in graduate students. A qualitative analysis]. *European Public & Social Innovation Review, 10*, 1–17.

Gallent, C.T., Zapata, A. G., & Ortego, J. L. H. (2023). El impacto de la inteligencia artificial generativa en educación superior: Una mirada desde la ética y la integridad académica. *RELIEVE: Revista Electrónica de Investigación y Evaluación Educativa.* https://doi.org/10.30827/relieve.v29i2.29134

Gustilo, L., Ong, E., & Lapinid, M. R. (2024). Algorithmically-driven writing and academic integrity: Exploring educators' practices, perceptions, and policies in the AI era. *International Journal for Educational Integrity, 20*(1), Article 3. https://doi.org/10.1007/s40979-024-00153-8

Hanlin, L., Yu, W., Siqi, L., & Changqing, H. (2024). The influence of GenAI on the effectiveness of argumentative writing in higher education: Evidence from a quasi-experimental study in China. *Journal of Asian Public Policy.* https://doi.org/10.1080/17516234.2024.2363128

Hazem, Z., John, M., & Mike, K. (2023). Ethics of generative AI. *Journal of Medical Ethics.* https://doi.org/10.1136/jme-2023-108909

Hind, A. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal.* https://doi.org/10.24093/awej/chatgpt.2

Imran, M., & Norah, A. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology.* https://doi.org/10.30935/cedtech/13605

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*. https://doi.org/10.1016/j.lindif.2023.102274.

King, C., Letlhogonolo Nolo, M., & Mfanelo Patrick, N. (2024). Challenges, benefits and recommendations for using generative artificial intelligence in academic writing: A case of ChatGPT. *Medicon Engineering Themes, 7*(6), 3–38. https://doi.org/10.31222/osf.io/7hr5v

Kleopatra, N. (2024). Generative artificial intelligence in higher education: Exploring ways of harnessing pedagogical practices with the assistance of ChatGPT. *International Journal of Curriculum and Education.* https://doi.org/10.47852/bonviewijce42022489

Konstantinos, T. K. (2025). *The role of ChatGPT in academic writing: Pedagogical and ethical dimensions.* https://doi.org/10.56294/ai2025420

Kovari, A. (2025). A systematic review of AI-powered collaborative learning in higher education: Trends and outcomes from the last decade. *Social Sciences & Humanities Open, 11*, 101335. https://doi.org/10.1016/j.ssaho.2025.101335

Leon, S., Ye, C., Ying, H. G., & Zhe, C. (2025). *Finding your voice: Using generative AI to help international students improve writing.* https://doi.org/10.20944/preprints202502.1524.v1

Mohammed Siddique, K. (2025). Written assignments and ethical considerations of artificial intelligence in higher education. *International Journal of Linguistics, Literature and Translation*, 8(9), Article 27. https://doi.org/10.32996/ijllt.2025.8.9.27

Mokhine-Martins, M., Deka, J., & Costa, K. (2024). Positive psychology interventions and initiatives – The Ubuntu approach – A mapping review protocol. *Journal of Research Methodologists, 3*(1). https://doi.org/10.63591/GCAR.JRM.3.1.76

Moorhouse, B. L., Wan, Y., Wu, C., Wu, M., & Ho, T. Y. (2025). Generative AI tools and empowerment in L2 academic writing. *System.* https://doi.org/10.1016/j.system.2025.103779

Muhammad, A., Aisha, A., Sadia, Q., Muhammad Sabboor, H., & Sadia, B. (2025). Academic writing issues in the AI era: Opportunities and challenges for applied and social sciences in Pakistan. *Academic Journal*, 4(3). https://doi.org/10.63056/acad.004.03.0583

Oates, A., & Johnson, D. (2025). ChatGPT in the classroom: Evaluating its role in fostering critical evaluation skills. *International Journal of Artificial Intelligence in Education.* https://doi.org/10.1007/s40593-024-00452-8

Promethi Das, D., & Yixin, C. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies.* https://doi.org/10.3390/soc15090247

Quynh Hoa, N. (2023). AI and plagiarism: Opinion from teachers, administrators and policymakers. *PAIC.* https://doi.org/10.54855/paic.2346

Romaric, D. (2024). ChatGPT has eaten my assignment: A student-centric experiment on supervising writing processes in the AI era. *EDUCON Proceedings.* https://doi.org/10.1109/educon60312.2024.10578917

Rosario Michel-Villarreal, E. L., Vilalta-Perdomo, E. L., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Flor Silvestre, G. (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Neveléstudomány.* https://doi.org/10.3390/educsci13090856

Saeed Awadh, B.-N., Mouad, S., & Mohamed, B. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society.* https://doi.org/10.1016/j.techsoc.2023.102370

Santosh, K., Patil, V. D., Parikshit, N. M., & Vishal, M. (2023). Concept and administration for developing special education using AI. In *Advances in Educational Technologies and Instructional Design* (Chapter 11). https://doi.org/10.4018/979-8-3693-0378-8.ch011

Sarin, S., & Kimkong, H. (2024). Opportunities, challenges, and strategies for using ChatGPT in higher education: A literature review. *Journal of Digital Educational Technology.* https://doi.org/10.30935/jdet/14027

Stefan, K., & Sebastian, W. (2023). Navigating the impact of ChatGPT/GPT-4 on legal academic examinations: Challenges, opportunities, and recommendations. *Media Iuris.* https://doi.org/10.20473/mi.v6i2.45270

Thaweesak, C. (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Edelweiss Applied Science and Technology.* https://doi.org/10.55214/25768484.v9i4.6128

Ting, W., Brady, D. L., Agostino, M., Pagano, A., Zoë, T., & Jenny, P. (2023). Exploring the potential impact of artificial intelligence (AI) on international students in higher education: Generative AI, chatbots, analytics, and international student success. *Applied Sciences.* https://doi.org/10.3390/app13116716

Tom, F., & Nick, B. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Neveléstudomány.* https://doi.org/10.3390/educsci13111109

Tony, B., Cristóbal, C., Olga, M., & Steve, W. (2020). Can artificial intelligence transform higher education? *International Journal of Educational Technology in Higher Education.* https://doi.org/10.1186/s41239-020-00218-x

Turingan, V. T. (2025). Exploring students' perspectives on utilization of artificial intelligence (AI) writing tools through sequential explanatory mixed method study. *International Journal of English Literature and Social Sciences, 10*(4). https://doi.org/10.22161/ijels.104.37

Wang, Y. D., Chen, J., Yuxin, W., Xi, Y., & Lixia, H. (2025). Impact of generative artificial intelligence on scientific paper writing and regulatory pathways. *Editing Practice.* https://doi.org/10.54844/ep2025.1011

Wojciech, W. (2023). ChatGPT in higher education. *International Journal of New Economics and Social Sciences.* https://doi.org/10.5604/01.3001.0053.9601

Xhulio, M., & Muhamet, Z. (2024). Using AI to improve accessibility and inclusivity in higher education for students with disabilities. https://doi.org/10.1109/ithet61869.2024.10837607

Xianghan, O. D. (2024). Generative AI: Is it a paradigm shift for higher education? *Studies in Higher Education.* https://doi.org/10.1080/03075079.2024.2332944

Xiaoli, X., & Sheng, D. (2023). Opportunities, challenges, strategies, and reforms for ChatGPT in higher education. https://doi.org/10.1109/eki61071.2023.00010

Yanhua, L., Jaeuk, P., & Sean, M. (2024). Using generative artificial intelligence/ChatGPT for academic communication: Students' perspectives. *International Journal of Applied Linguistics.* https://doi.org/10.1111/ijal.12574

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