IJEMST International Journal of Education in Mathematics, Science and Technology

www.ijemst.net

Generative Artificial Intelligence (GenAI) for Academic Writing in Higher Education: A Scoping Review of Applications, Challenges, and Implications

Renz Alvin E. Gabay [1], Aaron A. Funa [2] *, Jhonner D. Ricafort [3]

[1] Sorsogon State University, Sorsogon, Philippines, 0000-0002-8108-4589
[2] Sorsogon State University, Sorsogon, Philippines, 0000-0002-6648-8825
[3] Sorsogon State University, Sorsogon, Philippines, 0000-0001-8980-6681
* Corresponding author: Aaron A. Funa (funa.aaron@sorsu.edu.ph)

Article Info

Article History

Received:
19 August 2025

Revised:
30 October 2025

Accepted:
23 November 2025

Published:
1 January 2026

Keywords

Academic writing
ChatGPT
GAI
Higher education
Large language models
Scoping review

Abstract

Generative artificial intelligence (GenAI) is reshaping academic writing in higher education faster than institutions can develop evidence-informed guidance, leaving practice ahead of proof. To clarify what is happening and where benefits and risks cluster, the researchers conducted a scoping review structured by a Population–Concept–Context (PCC) frame and aligned with PRISMA-ScR procedures. Peer-reviewed, English-language empirical studies published from 2024 through Q2 2025 in higher-education settings were included, and findings were synthesized via convergent integration that juxtaposed quantitative distributions with qualitative themes. A total of 25 studies met criteria. Across populations and contexts, GenAI was most often positioned as assistive scaffolding across the planning-to-revision span of writing; reported benefits concentrated on organization, fluency, efficiency, and language support (notably for multilingual writers). Recurrent risks included hallucinations and unreliable or fabricated citations, inconsistent disclosure or attribution, and overreliance when use was unscaffolded; the limited reliability of AI-detection tools complicated integrity judgments. Context shaped practice: clearer policies and better access supported more constructive use, while the evidence base skews toward English-medium, well-resourced institutions and relies heavily on short-term or proxy outcomes. By integrating counts and themes within a PCC frame, this review offers an up-to-date evidence map that distinguishes where benefits reliably cluster (process-level supports) and where risks persist (source work and attribution), while surfacing salient gaps (faculty/postgraduate cohorts and Global South contexts). Overall, the pattern supports an assistive, not substitutive stance in which GenAI complements—rather than replaces—human judgment in argument construction, source interrogation, and synthesis.

Citation: Gabay, R. A. E., Funa, A. A., & Ricafort, J. D. (2026). Generative Artificial Intelligence (GenAI) for academic writing in higher education: A scoping review of applications, challenges, and implications. *International Journal of Education in Mathematics, Science and Technology (IJEMST), 14*(1), 200-232. https://doi.org/10.46328/ijemst.5682

Introduction

Generative artificial intelligence (GenAI; referred to as "GAI" in some studies) is transforming how people approach writing, communication, and knowledge creation. In the post-pandemic period, as remote learning and digital platforms became integral to education, tools such as ChatGPT, Gemini, and other large language models (LLMs) entered mainstream academic spaces and are now used for brainstorming, paraphrasing, and even drafting research papers (Emran et al., 2024; Meyer et al., 2023). Scholars, educators, and students increasingly recognize that these technologies can enhance productivity and creativity in academic writing (Funa & Gabay, 2025a; Khalifa & Albadawy, 2024). At the same time, their implications for academic integrity, authorship, and writing practice remain contested and not yet fully understood (Acut et al., 2024; Funa & Gabay, 2025b). This review examines the emerging literature on GenAI's role in academic writing, mapping current evidence, identifying gaps, and clarifying ongoing debates through a comprehensive scoping review. Understanding GenAI's influence is crucial for shaping future academic policies and pedagogies.

GenAI refers to AI systems designed to produce new content—text, images, or code—rather than merely analyze existing data. Its conceptual foundation can be traced to research on generative models, particularly the introduction of generative adversarial networks (GANs) by Goodfellow et al. (2014), which showed that AI could learn patterns in data and create novel outputs. Contemporary tools such as ChatGPT and Gemini are built on what the Stanford Center for Research on Foundation Models (CRFM) terms foundation models: large-scale, self-supervised deep learning systems trained on vast datasets and adaptable to many tasks. These models provide the underlying architectures that power modern GenAI applications and enable language, image, and code generation at scale (Bommasani et al., 2021).

In academic settings, GenAI tools are widely used for idea generation, paraphrasing, literature summarization, and drafting, reshaping how students and scholars approach writing (Acut et al., 2024; Emran et al., 2024; Funa & Gabay, 2025a; Kasneci et al., 2023; Meyer et al., 2023). A recent survey of medical students in the United States found that 48.9% had used ChatGPT in their studies; among users, 43.7% reported weekly to daily use, most commonly for writing, revising, editing, and summarizing. Notably, 37.5% and 41.3% reported using ChatGPT for these tasks for more than 25% of their working time (Zhang et al., 2024). Complementing higher education evidence, K–12 STEM classrooms likewise report perceived gains in interest, academic proficiency, and learning independence with ChatGPT (Diaz et al., 2025). Funa and Gabay (2025a) similarly observed that faculty across generations in higher education use GenAI primarily for ideation and rapid feedback, with younger participants tending to trust outputs more readily. Kasneci et al. (2023) highlight potential benefits for students and educators—such as quiz generation, simplification of complex content, and adaptive feedback—while cautioning about bias, overreliance, and ethical concerns. Consistent with these cautions, Funa and Gabay (2025b) and Meyer et al. (2023) emphasize that although GenAI can improve clarity, grammar, and readability, including for non-native English users, scholars should remain vigilant about factual inaccuracies, ethical issues, and model biases.

To interpret these emerging practices, this review draws on complementary theoretical frameworks. It is anchored in socio-constructivist perspectives on writing and learning, which view writing as a socially mediated process shaped by tools, contexts, and interaction. From this perspective, learners develop writing proficiency through dialogue with peers, engagement with cultural tools, and iterative practice supported by scaffolds (Flower & Hayes, 1981; Vygotsky, 1978). GenAI tools such as ChatGPT and Gemini can therefore be conceptualized as cognitive and metacognitive scaffolds that provide immediate feedback, offer alternative phrasings, and suggest structural improvements to support planning, revision, and refinement. These tools function as mediators in the social process of writing and may extend a writer's zone of proximal development by offering access to language models and ideas that might otherwise be unavailable. In parallel, the paradigm of foundation models (Bommasani et al., 2021) situates GenAI within a broader shift toward flexible, generalizable AI systems that mediate knowledge creation across disciplines. Integrating these perspectives provides a robust basis for examining how GenAI reshapes writing practices and for considering the pedagogical, ethical, and institutional implications that follow.

Against this backdrop, current scholarship on GenAI and academic writing continues to expand but often addresses isolated tools, single-discipline applications, or specific aspects of writing support. Comprehensive syntheses that map patterns of use across educational contexts remain limited (Kasneci et al., 2023). Early advances in LLMs began influencing educational technologies around 2020 and laid the groundwork for new AI-mediated learning tools. The public release of ChatGPT in November 2022 marked a turning point, prompting rapid growth in research and adoption, particularly as digital and AI-assisted platforms became integral to academic workflows in the post-pandemic period (Bisi et al., 2023; Bommasani et al., 2021; Huh, 2023; OpenAI, 2022).

This review adopts a scoping review approach, which is well suited to this topic because the field of GenAI is evolving rapidly, the available evidence is heterogeneous, and the objective is to map the breadth of existing literature rather than evaluate intervention effectiveness or test a narrowly defined hypothesis (Arksey & O'Malley, 2005; Tricco et al., 2018). It builds on a systematic review by Chanpradit (2025), who synthesized 30 empirical studies from 2023 to 2024 and reported gains in cohesion, clarity, creativity, fluency, and proficiency, alongside risks such as plagiarism, overreliance, hallucinations, bias, and unequal access. The review also recommended institutional guidelines, transparent data practices, human oversight, and structured training. Extending this work, the present scoping review covers studies published from 2024 through the second quarter of 2025 to capture both ongoing adoption and emerging developments in higher education. By synthesizing peer-reviewed studies from this period, the review maps current applications, challenges, and research gaps in the use of GenAI for academic writing and identifies priorities for further investigation.

Guided by the Population–Concept–Context (PCC) framework from the Joanna Briggs Institute (Arksey & O'Malley, 2005; Levac et al., 2010; Peters et al., 2020), the researchers define the population as individuals engaged in academic writing in higher education (students, faculty, and researchers), the concept as the use of GenAI in academic writing, and the context as higher-education settings. Accordingly, the researchers address: (RQ1) How is GenAI being used in academic writing? (RQ2) What benefits and opportunities are reported? (RQ3) What challenges, risks, or ethical issues are identified? (RQ4) What gaps and future research directions are highlighted by existing studies?

Methodology

Research Design

This scoping review followed the methodological framework of Arksey and O'Malley (2005), refined by Levac et al. (2010) and Peters et al. (2020), guided by procedures outlined in Funa et al. (2024), and reported in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR; Tricco et al., 2018). The approach was selected to map the breadth of evidence on the use of GenAI in academic writing and to identify research gaps in a rapidly evolving field. Eligibility and synthesis were structured using the PCC mnemonic: Population (students, faculty, and researchers in higher education), Concept (use of GenAI in academic writing), and Context (higher-education scholarly settings).

Search Strategy and Study Selection

An initial search was performed using Harzing's Publish or Perish (Harzing, 2007) to explore and retrieve relevant literature from multiple databases, including Google Scholar, Scopus, PubMed, Semantic Scholar, and Web of Science. A comprehensive list of keywords and descriptors was developed, refined, and iteratively tested across these databases. Boolean operators (AND, OR) were applied, and search terms were systematically combined and interchanged to ensure coverage across the three focal constructs: academic writing, GenAI, and higher education. For example, the primary term "academic writing" or "scientific writing" was combined with secondary terms such as "generative artificial intelligence," "generative AI," or "GenAI," together with "higher education" or "university students," using AND/OR.

Inclusion and Exclusion Criteria

Following the recommendations of Levac et al. (2010), the inclusion and exclusion criteria were designed to align directly with the research questions and were refined through iterative team discussions during the initial screening phase. The study selection process is shown in Figure 1. A study was included if it met all of the following conditions: (a) it was a peer-reviewed journal article published between 2024 and the second quarter of 2025, reflecting the rapid advancements and evolving applications of GenAI during this period; (b) it made an explicit reference to the use, application, or impact of GenAI—such as ChatGPT, Gemini, or other LLMs—within the context of academic writing/scientific writing; (c) it was written in English to ensure accurate interpretation; (d) it reported original empirical findings, ensuring the synthesis was based on primary data or firsthand analyses rather than secondary syntheses; and (e) it focused on higher education institutions. Conference papers and other non-journal sources were excluded to maintain the rigor and comparability of the included evidence. Restricting the review to empirical studies in higher education ensured that the findings provided robust, evidence-based insights into the phenomenon under investigation. The criteria were piloted and refined before the full screening process to enhance clarity and ensure consistent application.

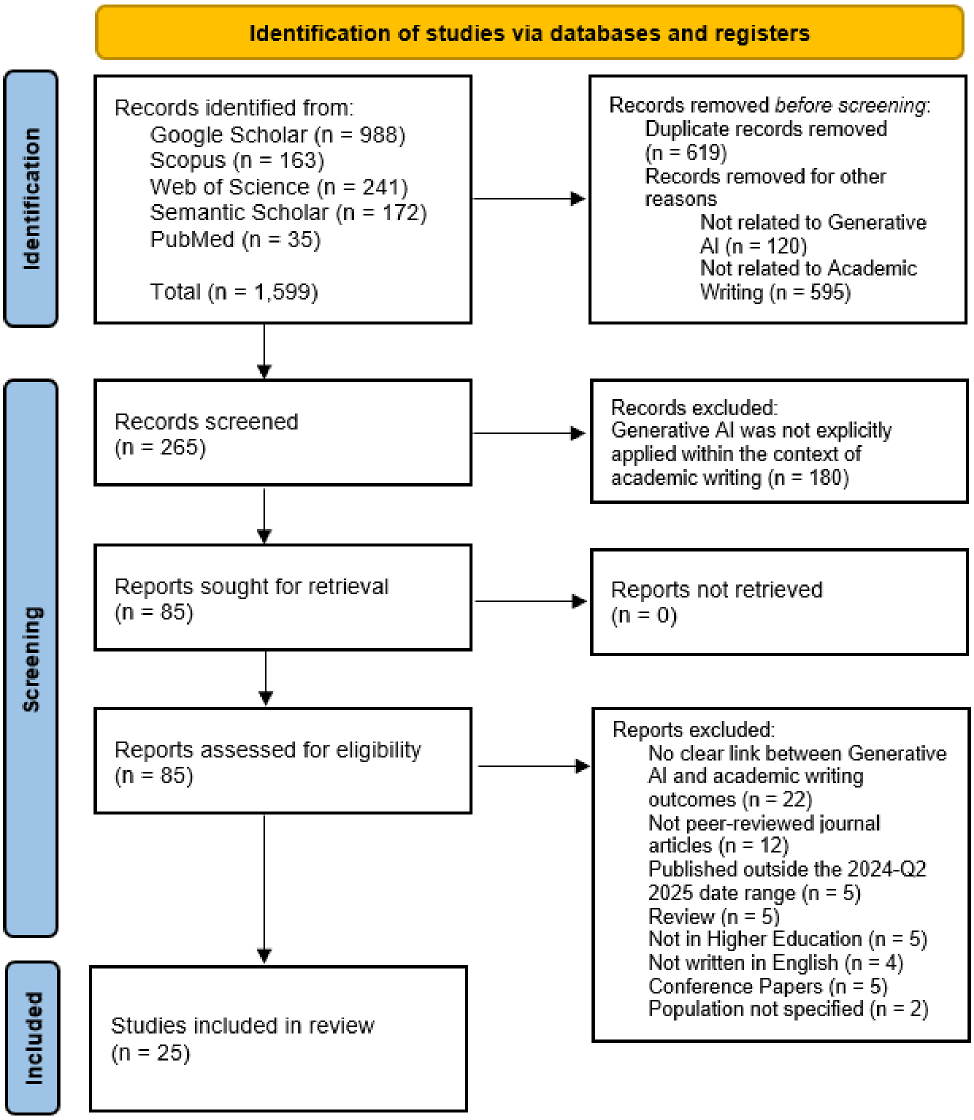

Figure 1. PRISMA Study Selection Flow Diagram for GenAI in Academic Writing

Figure 1 presents the PRISMA-ScR flow for study identification, screening, eligibility, and inclusion for this scoping review (2024–Q2 2025). Across databases, 1,599 records were identified; 1,334 were removed prior to screening, leaving 265 titles/abstracts screened. Of these, 180 were excluded at screening and 85 full texts were assessed for eligibility. 60 full-text articles were excluded (primary reasons summarized in Figure 1; the most common was no clear link between GenAI use and academic writing in higher education), yielding 25 studies included in the final synthesis.

Coding Procedures

Data from the included studies were extracted and organized using a standardized data charting form developed by the review team. In line with Levac et al. (2010), the coding process was iterative, allowing the team to refine categories as familiarity with the literature increased. Each article was coded for key variables such as publication details (author & year), country/region, population (students, faculty, researchers), study context, study design/methodology, type of GenAI tool/s used, specific applications in academic writing (e.g., brainstorming, drafting, paraphrasing, summarizing), reported benefits and opportunities, identified challenges and ethical issues, and stated research gaps or recommendations for future studies.

Two researchers independently coded each included study to enhance reliability; discrepancies were discussed and resolved through consensus, with a third researcher consulted when necessary. While independent coding supported organization, theme development relied on interpretive discussion rather than fixed reliability coefficients, consistent with reflexive thematic analysis by Braun & Clarke (2021). Excel was used to facilitate organization and thematic grouping of data. Codes were grouped into higher-order categories aligned with the research questions, and emerging themes were refined through ongoing team discussion. Consistent with scoping review guidance, the objective was to map the evidence; therefore, no formal critical appraisal of study quality was undertaken. Findings were synthesized using a descriptive numerical summary of study characteristics and a qualitative thematic synthesis mapped to the PCC framework and the review questions.

Characteristics of the Included Studies

This scoping review synthesized peer-reviewed empirical studies published from 2024 to Q2 2025 (n = 25) that investigated GenAI in academic writing within higher education contexts (see Appendix A). To aid interpretation, settings are grouped by continent (see Figure 2): Europe (n = 11: United Kingdom [n = 7], Denmark, Norway, Greece, Switzerland), Asia (n = 7: China [n = 3], Hong Kong [n = 2], United Arab Emirates, Türkiye), Americas (n = 4: Ecuador [n = 2], Chile, United States), and Africa (n = 1: South Africa), with two cross-regional studies (one bi-national Finland–New Zealand; one global/multi-country survey).

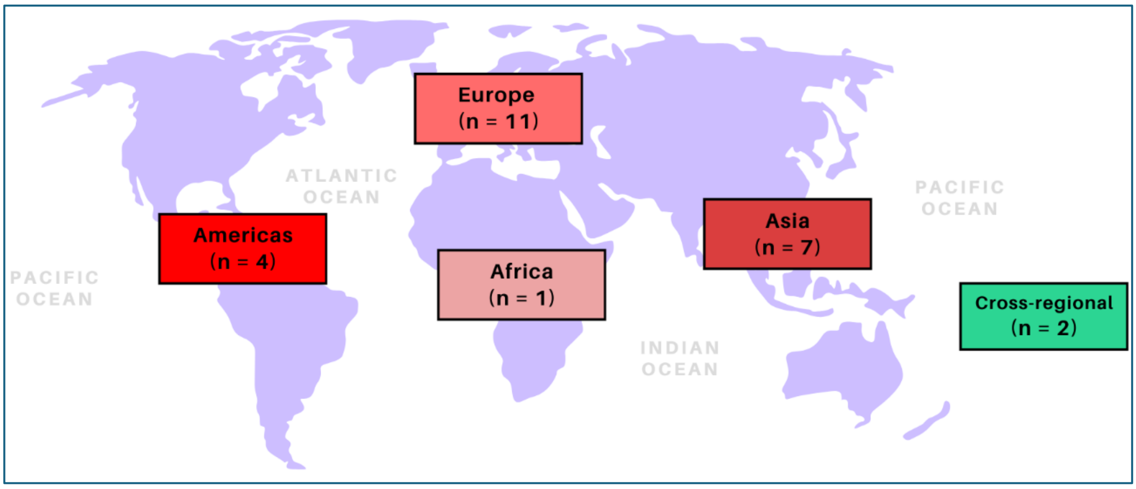

Figure 2. GenAI and Academic Writing Studies by Continent (2024–Q2 2025, n = 25).

Note: Türkiye counted under Asia.

Across 25 studies, participants were predominantly student-only samples: 19 of 25 (76%); faculty-only: 3 of 25 (12%); and mixed student-faculty: 3 of 25 (12%). Sample sizes ranged from small, course-embedded cohorts (e.g., 6 students in Denmark [Jensen & Jensen, 2025]; 20 students at a Sino-British EMI university [Kim et al., 2025]) to large cross-sectional surveys (e.g., 2,555 students at a UK university [Johnston et al., 2024]; a global mixed sample of n = 1,217 students and lecturers [Yusuf et al., 2024]). Staff-only samples also appeared (e.g., n = 284 UK academics [Watermeyer et al., 2024]; n = 184 university teachers in Ecuador [Cordero et al., 2025]).

Study designs were mainly descriptive and exploratory, comprising quantitative surveys (Johnston et al., 2024), qualitative interviews or focus groups (Hysaj et al., 2025), mixed-methods designs (Han, 2025), task-based observational studies (Johnston et al., 2025), and a small number of course-embedded interventions (Wang & Ren, 2024). The most commonly examined tools were ChatGPT (n = 23), followed by Grammarly (n = 4), Microsoft Copilot (n = 3), Perplexity (n = 2), QuillBot (n = 1), and local or institution-provided LLMs (n = 2). Reported writing applications centered on brainstorming/idea generation and planning/outlining (Nguyen et al., 2024; Johnston et al., 2025), as well as summarizing literature and drafting/refinement within course tasks (Wang & Ren, 2024). Across contexts, authors noted benefits such as efficiency/time savings and greater confidence/self-efficacy (Campbell & Cox, 2024; Hysaj et al., 2025), along with language support for L2 writers and organizational help (planning/structuring, grammar support) (Johnston et al., 2024). Recurrent challenges included risks of plagiarism/contract cheating and policy gaps/uneven AI literacy (Campbell & Cox, 2024; Hysaj et al., 2025), hallucinations and unverifiable sources (Johnston et al., 2025), overreliance that may erode writing skills and limitations of detection tools (Jensen & Jensen, 2025), and broader academic-integrity concerns (Nelson et al., 2025).

Data Analysis Procedures

Data analysis followed established scoping-review methodology that prioritizes mapping the breadth of evidence and key concepts over effect estimation (Arksey & O'Malley, 2005; Levac et al., 2010; Peters et al., 2020; Tricco et al., 2018). Guided by the PCC framework, the researchers conducted two complementary analytic strands and then integrated findings to address the review questions.

Using Microsoft Excel, the researchers obtained quantitative data by coding publication year, country and continent, participant group, study design, and GenAI tools. Multi-country items were coded as cross-regional. Tool tallies reflect studies that examined or instructed the use of a tool at least once; because some studies included multiple tools, tool counts can exceed n = 25. Cross-tabulations (for example, tool and participant group; design and continent) were generated to surface distributional patterns. Outputs included summary tables and figures (for example, PRISMA flow; continent map) consistent with PRISMA-ScR reporting (Tricco et al., 2018). Qualitative data extracted in the charting fields (applications in academic writing, benefits and opportunities, challenges and ethical issues, and gaps or recommendations) were analyzed using thematic analysis following Braun and Clarke's (2006) six-phase procedure. Braun and Clarke (2021) informed the reflexive stance and reporting. A deductive and inductive approach was applied: an a priori frame aligned to PCC and the review questions guided initial coding; line-by-line open coding allowed additional categories to emerge; a shared codebook was refined iteratively; and higher-order themes were developed through constant comparison and memoing. Quantitative and qualitative results were combined using a convergent integrated approach in the Discussion section: numerical patterns (for example, continent or tool frequencies) were juxtaposed with qualitative themes in side-by-side matrices and narrative weaving to produce an evidence map that directly answers each review question (Fetters et al., 2013; Peters et al., 2020).

Results

Findings are organized around the four research questions and integrate a descriptive numerical summary with a qualitative thematic synthesis. Drawing on the 25 included studies, the researchers first map how GenAI is being used in academic writing (RQ1), then synthesize reported benefits and opportunities (RQ2), followed by challenges, risks, and ethical issues (RQ3), and finally identify gaps and future research directions (RQ4). This structure aligns with the review's PCC framing and the stated research questions.

Ways GenAI is being used in Academic Writing (RQ1)

Across the 25 studies, GenAI is used across the full writing cycle, from pre-writing through finalization (see Table 1).

Table 1. Ways GenAI is used across Stages of Academic Writing (n = 25).

Writing stage	Typical GenAI uses	Representative tools	Example sources
Pre-writing	Clarify concepts, generate ideas/topics, produce examples	ChatGPT; local LLMs	Johnston et al. (2025); Johnston et al. (2024)
Planning & outlining	Create outlines, reorganize structure, plan sections	ChatGPT; Copilot; Perplexity	Johnston et al. (2025); Nguyen et al. (2024)
Drafting	Produce first drafts, expand points, suggest wording	ChatGPT	Wang & Ren (2024)
Revising & refining	Rewrite passages, improve coherence/flow, style and tone	ChatGPT	Wang & Ren (2024); Hysaj et al. (2025)
Language support (L2)	Paraphrase/translate, vocabulary support, grammar/mechanics	Grammarly; QuillBot; ChatGPT	Hysaj et al. (2025); Johnston et al. (2024)
Working with sources	Generate search terms, summarize articles, format references	Perplexity; ChatGPT; Copilot	Johnston et al. (2025); Jensen & Jensen (2025)
Multimodal support	Create/plan visuals to accompany text; slide notes	ChatGPT (image features)	Wang & Ren (2024)

Students most commonly employ chatbots to clarify concepts, request definitions and examples, and generate outlines or essay plans, indicating early-stage support for planning and structuring (Johnston et al., 2025). GenAI is then used to draft and reorganize text, polish language, and refine coherence, with classroom data showing perceived utility for idea generation, vocabulary support, grammar correction, and argument organization (Wang & Ren, 2024). In multilingual contexts, learners report using GenAI to paraphrase, simplify readings, and address language mechanics, describing these uses as practical supports for completing written assessments (Hysaj et al., 2025). Source-work and referencing support also appear frequently, although students' practices are uneven; for example, students plan and search with GenAI but often omit acknowledging the tool itself in references, underscoring an ongoing need for information-literacy guidance (Johnston et al., 2025; Jensen & Jensen, 2025). ChatGPT is the most frequently used tool, with additional use of Grammarly, Copilot, Perplexity, QuillBot, and institution-provided or local LLMs in several studies (Johnston et al., 2025).

As shown in Table 1, the studies depict whole-process integration rather than single-point use. Planning and conceptual scaffolding are especially prominent, followed by text production and language polishing. Source-work support is common, but attribution and referencing practices lag; students frequently use GenAI when planning or searching yet omit citing the tool, indicating a gap for AI-informed information-literacy instruction (Johnston et al., 2025; Jensen & Jensen, 2025). While ChatGPT dominates use, several studies encourage broadening tool awareness to alternatives such as Copilot and Perplexity, and to institution-provided LLMs where available (Johnston et al., 2025).

Benefits and Opportunities Reported regarding GenAI-assisted Writing (RQ2)

As summarized in Figure 3, the researchers identified six, recurrent benefit clusters that map onto the writing process (planning, drafting, revising, and finalizing). First, students consistently reported efficiency and time savings, with GenAI handling lower-level mechanics (grammar, phrasing, formatting) and thereby freeing attention for higher-order concerns such as argumentation and evidence use (Campbell & Cox, 2024; Han, 2025; Wang & Ren, 2024). Second, GenAI offered substantial language support, particularly in EAP/L2 contexts, where learners used it to paraphrase, translate, expand vocabulary, and improve clarity and fluency; these functions were frequently linked to increased confidence and self-efficacy in completing written assessments (Hysaj et al., 2025; Moorhouse et al., 2025; Nelson et al., 2025).

Figure 3. Benefits and Opportunities of GenAI-assisted Academic Writing (2024–Q2 2025; n = 25).
Note. Themes synthesized from the included studies and mapped to writing stages.

Third, tools were widely used for planning and organization—brainstorming, outlining, and structuring paragraphs—often improving task interpretation and the perceived coherence of drafts (Johnston et al., 2025; Mo & Crosthwaite, 2025; Wang & Ren, 2024). Fourth, several studies highlighted formative feedback and scaffolding: rapid explanations, exemplars, and revision suggestions supported iterative improvement; in some cases, source-display features helped students plan searches and check claims (Jensen & Jensen, 2025; Johnston et al., 2025).

Fifth, the literature points to access and inclusion opportunities. Students with disabilities described GenAI as helpful for planning, drafting, and multimodal expression (e.g., generating alternatives or simplifying language), indicating potential to reduce participation barriers when used with appropriate guidance (Zhao et al., 2025). Finally, at the pedagogical and institutional levels, studies framed GenAI as an opportunity space for course/assessment redesign and AI-literacy development—for example, integrating transparent AI use into authentic assessment, and embedding guidance on prompting, verification, and attribution (Cordero et al., 2025; Johnston et al., 2024; Kofinas et al., 2025).

Taken together, the pattern across contexts suggests GenAI's strongest contributions cluster in the planning-to-polishing span of the writing cycle. Benefits are maximized when use is transparent, scaffolded, and paired with verification practices, positioning GenAI as an assistive resource rather than a substitute for disciplinary thinking and academic integrity (Jensen & Jensen, 2025; Johnston et al., 2025).

Challenges, Risks, and Ethical Issues identified in the Literature (RQ3)

Synthesizing the 25 studies, the researchers identified five recurrent risk domains that cut across the writing process (see Figure 4). First, source reliability and epistemic risk remain prominent. Studies document hallucinations, unverifiable claims, and fabricated citations that can be difficult for novice writers to detect, underscoring the need for systematic verification and triangulation when GenAI is used for content generation or source work (Johnston et al., 2025; Jensen & Jensen, 2025; Zizka, 2025). Second, issues of authorship, attribution, and academic integrity are widely reported. Surveys and classroom investigations note inconsistent acknowledgment of tool use, uncertainty about the boundary between acceptable support and misconduct, and risks of plagiarism or contract cheating when outputs are submitted with minimal transformation (Johnston et al., 2024; Kofinas et al., 2025; Nelson et al., 2025; Johnston et al., 2025). Third, the literature points to overreliance and skill development concerns. Without scaffolding, students may defer critical reading, argumentation, and revision to GenAI, exhibiting automation bias and reduced practice with higher-order writing skills; several papers caution that "hands-off" use can erode competence over time (Jensen & Jensen, 2025; Han, 2025; Watermeyer et al., 2024). Fourth, assessment alignment and detection limits pose persistent challenges. Tasks that can be solved by generic prompts invite superficial engagement, while AI-detection tools are unreliable, susceptible to manipulation, and not suitable as sole evidence in integrity processes. The literature instead recommends process-focused assessment, multi-source evidence in investigations, and explicit expectations about permissible GenAI support (Kofinas et al., 2025; van Niekerk et al., 2025; Jensen & Jensen, 2025). Fifth, studies surface governance, equity, and privacy issues. Uneven AI literacy and unclear policies produce inconsistent practice across courses;

students also express concerns about sharing assignments with third-party systems, potential data exposure, and unequal access to paid tools and connectivity that can amplify existing inequities (Campbell & Cox, 2024; Johnston et al., 2024; Rodafinos, 2025; Zhao et al., 2025; Stanford, 2025).

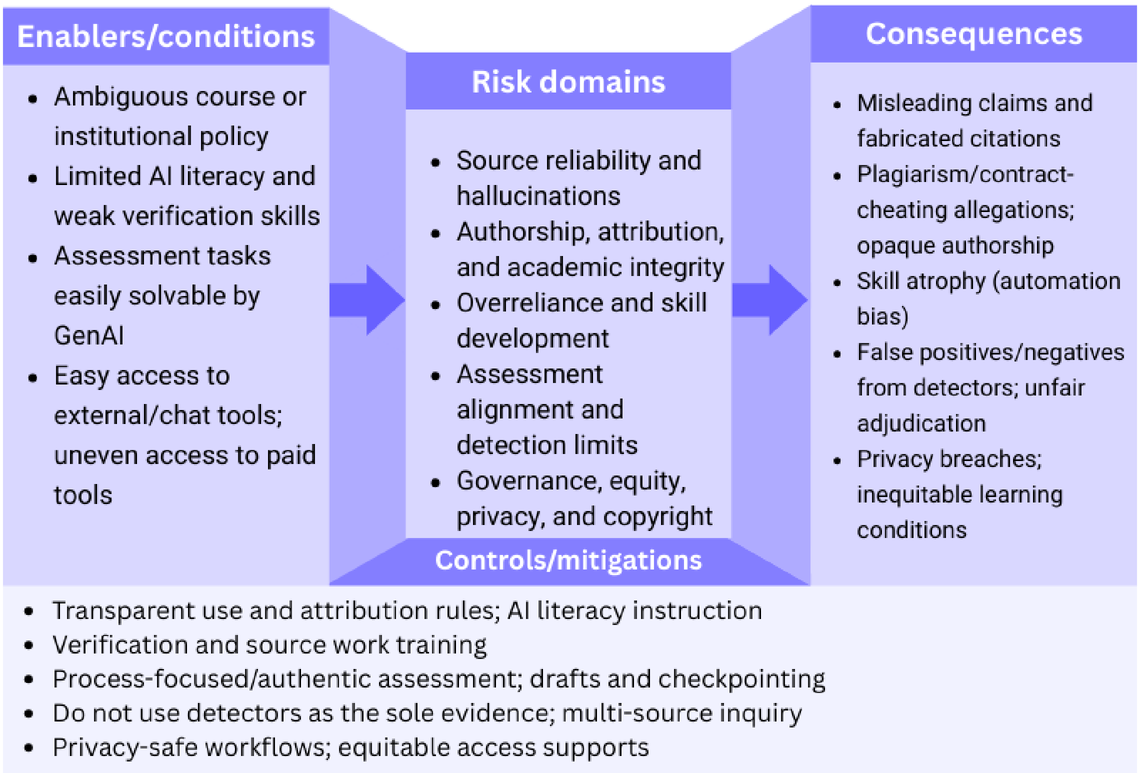

Figure 4. Bow-tie Map of Challenges, Risks, and Ethical Issues in GenAI-assisted Academic Writing (2024–Q2 2025; n = 25)

Note. Items synthesize findings across the included studies; arrows indicate progression from enabling conditions to risk domains and consequences. Controls/mitigations summarize commonly recommended responses.

The synthesized studies suggest that risk is highest where GenAI intersects with source credibility, authorship norms, and assessment design. Consistent recommendations across studies include transparent policies with required disclosure of tool use, explicit training in verification and source work, redesign of assessments to emphasize process and originality (e.g., staged drafts and checkpointing), avoidance of detectors as the sole basis for misconduct decisions, and privacy-safe, equitable access practices (Kofinas et al., 2025; Johnston et al., 2024; van Niekerk et al., 2025).

Gaps and Future Research Directions highlighted by Synthesized Studies (RQ4)

The researchers distilled seven cross-cutting gaps and aligned each with suggested methods and target outcomes (see Table 2). Collectively, the evidence base remains largely descriptive; stronger causal designs are needed to test whether GenAI improves writing quality, learning, or transfer across time and genres. Priority studies include longitudinal cohorts and classroom experiments with transparent comparison conditions and validated rubrics

(Kofinas et al., 2025; Wang & Ren, 2024; Kim et al., 2025; Moorhouse et al., 2025; Zizka, 2025).

Table 2. Summary of Research Gaps, Suggested Methods, and Target Outcomes in GenAI-Assisted Academic Writing (n = 25)

Research gaps (what needs evidence)	**Suggested methods (how to study it)**	**Target outcomes to report (what to measure)**	**Representative studies**
1. Causal effects on writing quality, learning, and transfer	Longitudinal cohorts; classroom experiments or quasi-experiments with transparent comparison conditions; multi-site replications	Writing quality using validated rubrics; learning gains and transfer to new genres; persistence over time; time-on-task and revision productivity	Kofinas et al., 2025; Wang & Ren, 2024; Kim et al., 2025; Moorhouse et al., 2025; Zizka, 2025
2. Reporting and attribution practices	Field experiments embedding disclosure requirements; audit studies of assignments; mixed-methods studies of student/marker perceptions	Disclosure rates; accuracy/completeness of AI-use statements; impacts on grades and feedback; perceived fairness/integrity	Johnston et al., 2025; Jensen & Jensen, 2025; van Niekerk et al., 2025; Johnston et al., 2024
3. AI-literacy and pedagogy interventions	Design-based research (DBR) on curricula; randomized or quasi-experimental evaluation of modules on prompting, verification, and source work	AI-literacy competency gains; verification accuracy; quality of source work; metacognitive strategy use; student confidence/self-efficacy	Campbell & Cox, 2024; Mo & Crosthwaite, 2025; Hysaj et al., 2025; Nguyen et al., 2024
4. Assessment and policy design	Comparative studies of assessment formats (staged drafts, in-class writing, viva/checkpoints); policy implementation evaluations; process tracing of drafting workflows	Misconduct allegations and outcomes; detector false-positive/negative rates; policy compliance/fidelity; marker workload; student satisfaction and perceived fairness	Kofinas et al., 2025; van Niekerk et al., 2025; Johnston et al., 2024
5. Equity, accessibility, and privacy	Studies with disability subgroups; usability testing; surveys/interviews on access; privacy impact assessments	Accessibility gains; accommodation effectiveness; access gaps (device, bandwidth, paid tools); privacy incidents and data sharing; differential outcomes by subgroup	Zhao et al., 2025; Watermeyer et al., 2024; Rodafinos, 2025; Stanford, 2025
6. Multilingual and cross-cultural contexts	Cross-site studies in non-Anglophone HEIs; L1/L2 comparisons; corpus-informed analyses of genre and register	Language quality (clarity, cohesion, accuracy) in L2 writing; translation/paraphrase fidelity; genre conformity; cultural/disciplinary fit	Nelson et al., 2025; Adalı & Bilgili, 2025; Hysaj et al., 2025
7. Measurement standards and replication	Consensus methods (e.g., Delphi) to define taxonomies; preregistered protocols; shared prompts and model versions; open materials for replication	Reporting-checklist compliance; reproducibility of results; sensitivity to model/version/prompt; open datasets and code availability	Mo & Crosthwaite, 2025; Han, 2025

A second cluster concerns reporting and attribution. Multiple papers document inconsistent disclosure of GenAI use and uncertainty about acknowledgment norms, especially when tools assist with planning or paraphrase. Table 3 therefore recommends field experiments that embed disclosure requirements and mixed-methods audits of student and marker perceptions, with outcomes such as disclosure rates, accuracy of AI-use statements, and effects on grading and feedback (Johnston et al., 2025; Jensen & Jensen, 2025; van Niekerk et al., 2025; Johnston et al., 2024). Third, the literature calls for systematic evaluation of AI-literacy and pedagogy. While many studies advocate instruction in prompting, verification, and source work, few evaluate structured curricula at scale. Design-based research and quasi-experimental modules should report competency gains, verification accuracy, quality of source work, and metacognitive strategy use (Campbell & Cox, 2024; Mo & Crosthwaite, 2025; Hysaj et al., 2025; Nguyen et al., 2024).

Fourth, assessment and policy design require rigorous testing. Recommended directions include comparative evaluations of process-oriented formats (e.g., staged drafts, in-class writing, oral checkpoints), policy implementation studies, and process tracing of drafting workflows; outcomes should include integrity incidents, false-positive rates from detectors, policy fidelity, marker workload, and perceived fairness (Kofinas et al., 2025; van Niekerk et al., 2025; Johnston et al., 2024). Fifth, equity, accessibility, and privacy remain under-researched. Early findings suggest potential to reduce participation barriers for students with disabilities but also raise concerns about unequal access to paid tools and data-sharing risks. Future work should incorporate subgroup analyses, usability testing, and privacy-impact assessments with clear reporting of accessibility gains and differential outcomes (Zhao et al., 2025; Watermeyer et al., 2024; Rodafinos, 2025; Stanford, 2025).

Sixth, the field needs broader coverage of multilingual and cross-cultural contexts. Studies should move beyond Anglophone settings to examine L1/L2 differences, translation/paraphrase fidelity, and genre conformity in non-English HEIs (Nelson et al., 2025; Karahan Adalı & Bilgili, 2025; Hysaj et al., 2025). Finally, measurement standards and replication would improve comparability. The researchers recommend consensus on use-case taxonomies (planning, drafting, revising), preregistered protocols, and sharing of prompts, model versions, datasets, and code; core outcomes should include rubric-based writing quality, learning/transfer, integrity outcomes, and sensitivity to model/version changes (Mo & Crosthwaite, 2025; Han, 2025).

Discussion

The researchers interpreted the map of evidence through the PCC lens using convergent integration, aligning numerical distributions with qualitative themes to explain what the patterns mean for academic writing in higher education rather than restating procedures or counts (Fetters et al., 2013; Tricco et al., 2018). As depicted in Figure 5, a PCC-aligned convergent model guides this interpretation: Population, Concept, and Context flow into a convergent-integration step (shown as a dashed band) that positions GenAI as assistive scaffolding while core scholarly practices remain grounded in human judgment. This process emphasis aligns with evidence that explicit argument scaffolds measurably improve conceptual understanding across levels and delivery modes, indicating that GenAI should be paired with argument-based routines rather than replace them (Ramallosa et al., 2022).

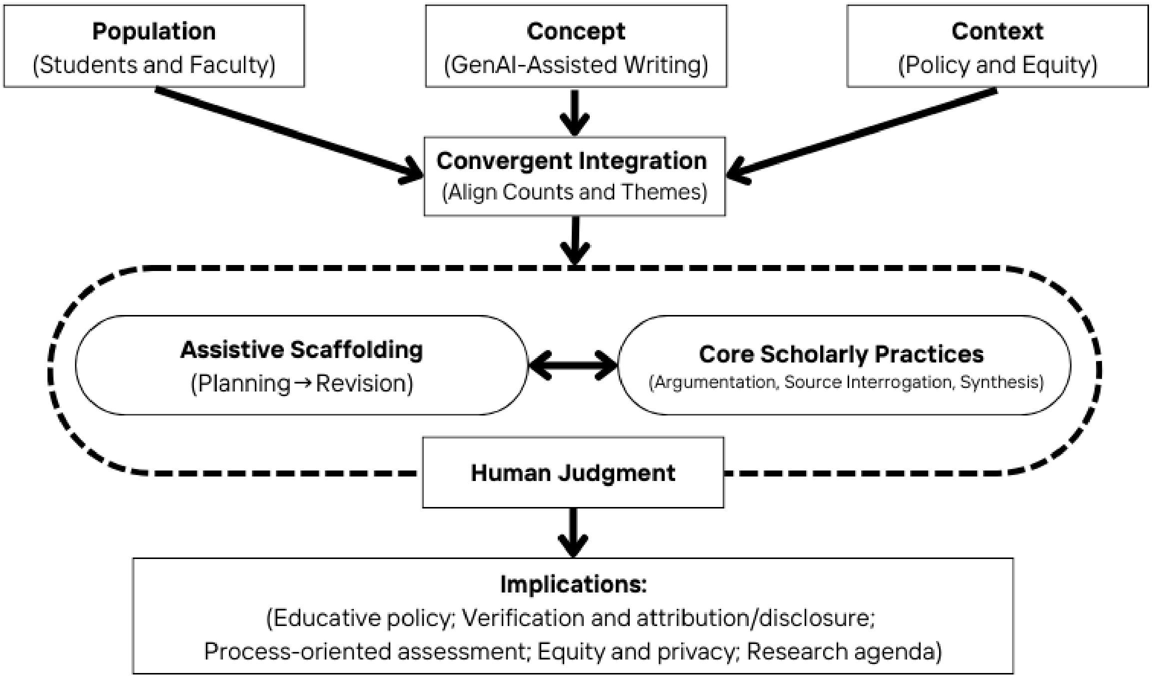

Figure 5. PCC-aligned Convergent Integration Model for GenAI in Academic Writing

With respect to RQ1, the researchers found convergence between tool prevalence and narratives of practice: widespread reliance on general-purpose chatbots alongside grammar and paraphrase tools corresponds to reported gains in clarity, organization, and confidence, especially among multilingual writers. Yet, qualitative accounts also underline that idea development, disciplinary reasoning, and source-critical reading are not reliably automated, which situates GenAI as a facilitator of process rather than a substitute for scholarly authorship (Jensen & Jensen, 2025; Johnston et al., 2025). This emphasis on process support is represented on the left of Figure 5 as “Assistive scaffolding (Planning to Revision).”

For RQ2 and RQ3, the researchers observed a stable pattern of benefits—efficiency, fluency, and reduced language barriers—counterbalanced by recurrent risks, notably hallucinations, fragile or fabricated citations, uneven disclosure, and signs of overreliance where scaffolding is absent. Divergence between stakeholder perspectives is salient: students often report improved fluency and task confidence, whereas instructors more frequently raise concerns about originality, source quality, and process visibility. The synthesis therefore supports the view that GenAI should augment, not replace, disciplinary thinking and source work, and that verification practices must be made explicit within assignments and feedback cycles (Campbell & Cox, 2024; Johnston et al., 2024; Zizka, 2025). The central “Human judgment” node in Figure 5 visually anchors this requirement for verification and attribution/disclosure. Positioning GenAI as a tool that supports claim–evidence–reasoning and structured rebuttal is consistent with meta-analytic gains from argument-based learning (Ramallosa et al., 2022), while keeping human verification central.

Interpreting the evidence through PCC also highlights where convergence and divergence matter. On population, the student-heavy corpus limits insight into supervisory practices, disclosure norms, and assessment decisions among faculty and postgraduate researchers. The researchers therefore note the need for more staff-focused and mixed-cohort studies to illuminate how expectations are translated into grading and feedback in authentic settings (Watermeyer et al., 2024; Cordero et al., 2025). On concept, whole-process assistance rather than single-step substitution explains the robust improvements in organization and language contrasted with mixed results for higher-order reasoning (Jensen & Jensen, 2025). On context, institutions with educative, transparent policies report more constructive uses—such as declared assistance with verification—while resource-constrained settings surface equity, access, and privacy concerns more sharply, suggesting that policy effectiveness is contingent on local conditions and support (Moorhouse et al., 2025; Zhao et al., 2025). Findings from secondary science during distance education show that learner-centered, action-oriented, and transformative practices emerged despite constraints, but inadequate equipment and poor connectivity were persistent barriers (Funa et al., 2023); these realities should shape GenAI policy to avoid deepening access gaps. These PCC elements are shown at the top of Figure 5, with directional arrows into the convergent-integration band.

These patterns help resolve the assistive–substitution debate. Where courses embed GenAI as taught scaffolding within a staged writing process, the researchers observed improvements without systematic loss of authorial voice. Where use is broad and unscaffolded, automation bias, shallow source engagement, and dependence are more likely. The literature converges on three guardrails: verifiability of content and references, transparent attribution or disclosure, and process-oriented assessment that makes thinking visible; by contrast, sole reliance on AI-detection tools is widely considered insufficient evidence for adjudicating integrity (Han, 2025; Kofinas et al., 2025; van Niekerk et al., 2025). Converging with these findings, a recent meta-analysis of inquiry-based learning likewise reports substantial improvements in students' conceptual understanding, particularly under open-inquiry conditions (Mediana Jr. et al., 2025); by analogy, GenAI should be embedded as a scaffold that sustains inquiry and verification rather than as an autonomous text generator. The bidirectional arrow between the two teal panels in Figure 5 signals interaction—GenAI can support process without supplanting core scholarly practices.

The researchers' implications also resonate with the I-STEM-PBL-ESD instructional framework, which integrates problem-based learning with education for sustainable development; GenAI tasks can be embedded as supports for problem framing, scenario exploration, and reflective synthesis while safeguarding attribution and process visibility (Funa et al., 2024). Implications follow directly for curriculum and policy. The researchers interpret the evidence as supporting AI-literacy that teaches prompting for thinking, verification strategies, and explicit attribution norms; assessment designs that document process via staged drafts, in-class checkpoints, and brief oral or written justifications; and equity-minded implementation that addresses differential access, accessibility needs, and data-privacy risks. These directions align with emerging institutional scholarship that emphasizes responsible, transparent adoption calibrated to local contexts (Funa & Gabay, 2025a, 2025b; Cordero et al., 2025).

Positioning this review within prior syntheses, the researchers corroborate early findings of clarity and fluency gains alongside integrity and equity concerns, while extending the field by focusing on 2024 to Q2 2025 studies and by integrating numerical distributions with qualitative themes. The added value lies in specifying where benefits reliably cluster (planning-to-polishing stages), which risks remain unresolved (verification and attribution), and which methodological moves are now necessary to advance knowledge—namely causal and longitudinal designs testing learning and transfer, systematic studies of disclosure and authorship practices, cross-cultural comparisons beyond Anglophone/EMI contexts, and clearer reporting standards for AI involvement (Chanpradit, 2025; Mo & Crosthwaite, 2025; Tricco et al., 2018). Consistent with Figure 5, the implications panel synthesizes these priorities: educative policy; verification and attribution/disclosure; process-oriented assessment; equity and privacy; and a forward research agenda.

Overall, the researchers interpret GenAI as most effective when treated as an assistive, transparent scaffold embedded in educative policy and process-forward pedagogy. Under these conditions, gains in organization, clarity, and productivity are achievable without compromising authorial ownership; under substitution framings, the risks noted above intensify. The field therefore benefits from designs and policies that preserve human judgment while leveraging GenAI's strengths in planning, feedback, and language support (Jensen & Jensen, 2025; Johnston et al., 2025; Zhao et al., 2025).

Limitations of the Study

The researchers conducted a scoping review to map GenAI use in higher-education writing, not to judge study quality or estimate effects. Inclusion filters (peer-reviewed, English, higher education, 2024–Q2 2025) may introduce language/publication bias and exclude preprints and conference work. Rapidly evolving terminology and tool stacks mean some relevant studies may have been missed, and conclusions rest on primary studies that often use small, single-site samples and proxy outcomes. The evidence base is skewed toward better-resourced, English-medium institutions with limited faculty/postgraduate and Global South representation. Finally, results may shift as GenAI capabilities and institutional policies change, and theme coding involved researcher judgment.

Conclusions and Recommendations

The researchers conclude that GenAI currently functions most productively as assistive scaffolding across the planning-to-revision span of academic writing in higher education. When paired with explicit expectations for verification, attribution or disclosure, and visible processes, GenAI can improve organization, clarity, and productivity without displacing core scholarly practices such as argument construction, source interrogation, and synthesis. Divergences in outcomes appear when use is unscaffolded, when originality is treated as a product metric rather than a process, and when institutions rely on detection tools as sole evidence of integrity. The present map also shows important evidence gaps, including limited representation of faculty and postgraduate cohorts, under-representation of the Global South, and an over-reliance on short-term proxy measures of learning. Taken together, these patterns support an assistive-not-substitutive stance that centers human judgment while leveraging GenAI for formative support.

The researchers recommend an institution-wide approach that treats GenAI as assistive scaffolding within clearly articulated policies and taught practices. Specifically, universities should implement AI-literacy programs that teach prompting for thinking, verification strategies, and explicit attribution or disclosure; course policies should define permitted uses, privacy expectations, and data handling; and faculty development should emphasize task designs that keep human reasoning visible and feedback that targets argument quality and source use. Assessment should be process-centered—requiring staged drafts, in-class checkpoints, and brief oral or written justifications—to make authorship and decision pathways transparent; institutions should avoid sole reliance on AI-detection tools and instead use multi-source evidence that includes process artifacts and instructor judgment. To promote equity and access, the researchers advise providing institutionally vetted tools, accessibility features, and low-bandwidth or offline pathways so that GenAI support does not widen existing gaps. In course design, GenAI should scaffold question generation, planning, formative feedback, and reflective synthesis while reserving claim-evidence-reasoning, source evaluation, and final interpretations for learners. Finally, the research agenda should prioritize causal and longitudinal studies of learning and transfer, broaden representation to include faculty and postgraduate cohorts and non-Anglophone or resource-constrained contexts, and improve reporting standards by documenting model versions, prompts, guardrails, and disclosure practices while employing outcomes that capture reasoning, ethics, and sustained learning rather than only surface features.

References

Acut, D. P., Malabago, N. K., Malicoban, E. V., Galamiton, N. S., & Garcia, M. B. (2024). "ChatGPT 4.0 Ghosted Us While Conducting Literature Search:" Modeling the Chatbot's Generated Non-Existent References Using Regression Analysis. *Internet Reference Services Quarterly*, *29*(1), 27–54. https://doi.org/10.1080/10875301.2024.2426793

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Bisi, T., Risser, A., Clavert, P., Migaud, H., & Dartus, J. (2023). What is the rate of text generated by artificial intelligence over a year of publication in Orthopedics & Traumatology: Surgery & Research? Analysis of 425 articles before versus after the launch of ChatGPT in November 2022. *Orthopaedics & Traumatology: Surgery & Research*, *109*(8), 103694.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, V., Card, D., Castellon, D., Chen, A., Creel, K. A., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., … Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv.* https://doi.org/10.48550/arXiv.2108.07258

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

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.

Campbell, L. O., & Cox, T. D. (2024). Utilizing generative AI in higher education teaching and learning. *Journal of the Scholarship of Teaching and Learning, 24*(4), 162–173. https://doi.org/10.14434/josotl.v24i4.36575

Chanpradit, T. (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Edelweiss Applied Science and Technology, 9*(4), 889–906. https://doi.org/10.55214/25768484.v9i4.6128

Cordero, J., Torres-Zambrano, J., & Cordero-Castillo, A. (2025). Integration of generative artificial intelligence in higher education: Best practices. *Education Sciences, 15*(1), 32. https://doi.org/10.3390/educsci15010032

Diaz, D., Funa, A. A., & Gabay, R. A. E. (2025). ChatGPT in STEM classrooms: Students' perceptions of interest, academic proficiency, and learning independence. *Journal of Basic Education Research, 6*(3), 470–480. https://doi.org/10.37251/jber.v6i3.2096

Emran, A. Q., Mohammed, M. N., Saeed, H., Keir, M. Y. A., Alani, Z. N., & Ibrahim, F. M. (2024, January). Paraphrasing ChatGPT answers as a tool to enhance university students' academic writing skills. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 501-505). IEEE.

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research, 48*(6 Pt 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600

Funa, A. A., & Gabay, R. A. E. (2025a). *Exploring perspectives toward artificial intelligence integration in science education: A cross-generational study* [Preprint]. SSRN. https://doi.org/10.2139/ssrn.5240513

Funa, A. A., & Gabay, R. A. E. (2025b). Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study. *Social Sciences & Humanities Open*, *11*, 101221.

Funa, A. A., Gabay, R. A. E., Esdicul, K. L., & Prudente, M. S. (2023). Secondary teachers' and students' perceptions of distance education in science: Focus on learner-centered, action-oriented, and transformative learning. *Dalat University Journal of Science, 13*(3), 156–181. https://doi.org/10.37569/DalatUniversity.13.3.1108(2023)

Funa, A. A., Ricafort, J. D., Jetomo, F. G. J., & Lasala Jr, N. L. (2024). Effectiveness of brain-based learning toward improving students' conceptual understanding: a meta-analysis. *International Journal of Instruction*, *17*(1), 361-380.

Funa, A. A., Roleda, L. S., & Prudente, M. S. (2024). Integrated Science, Technology, Engineering, and Mathematics—Problem-Based Learning—Education for Sustainable Development (I-STEM-PBL-ESD) framework. In D. Luberto, G. Sharma, N. Capraro, & R. Yadav (Eds.), *A Diversity of Pathways Through Science Education*. Springer. https://doi.org/10.1007/978-981-97-2607-3_9

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Han, Y. (2025). Beyond the algorithm: Reconciling generative AI and human agency in academic writing education. *International Journal of Learning and Teaching, 11*(1), 39–42. https://doi.org/10.18178/ijlt.11.1.39-42

Harzing, A. W. (2007). *Publish or Perish* [Computer software]. https://harzing.com/resources/publish-or-perish

Huh, S. (2023). Emergence of the metaverse and ChatGPT in journal publishing after the COVID-19 pandemic. *Science Editing*, *10*(1), 1-4.

Hysaj, A., Dean, B. A., & Freeman, M. (2025). Exploring the purposes and uses of generative artificial intelligence tools in academic writing for multicultural students. *Higher Education Research & Development*. Advance online publication. https://doi.org/10.1080/07294360.2025.2488862

Jensen, T. W., & Jensen, S. W. (2025). AI literacy in the context of working with sources: Pitfalls and possibilities of generative AI models in academic writing. *Journal of Academic Writing, 15*(S2), 1–12.

https://doi.org/10.18552/joaw.v15iS2.1224

Johnston, H., Eaton, M., Henry, I., Deeley, E.-M., & Parsons, B. N. (2025). Discovering how students use generative artificial intelligence tools for academic writing purposes. *Journal of Learning Development in Higher Education*, (34).

Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity, 20*, Article 2. https://doi.org/10.1007/s40979-024-00149-4

Karahan Adalı, G., & Bilgili, A. (2025). Generative AI in higher education: Students' perspectives on adoption, ethical concerns, and academic impact. *Acta Infologica, 9*(1), 147–166. https://doi.org/10.26650/acin.1670197

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Peters, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, *5*, 100145.

Kim, J., Lee, S.-S., Detrick, R., Wang, J., & Li, N. (2025). Students–generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*. Advance online publication. https://doi.org/10.1007/s12528-025-09444-6

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies, 30*, 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Kofinas, A. K., Tsay, C.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology, 00*, 1–28. https://doi.org/10.1111/bjet.13585

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science, 5*(69). https://doi.org/10.1186/1748-5908-5-69

Maphoto, K. B., Sevnarayan, K., Mohale, N. E., Suliman, Z., Ntsopi, T. J., & Mokoena, D. (2024). Advancing students' academic excellence in distance education: Exploring the potential of generative AI integration to improve academic writing skills. *Open Praxis, 16*(2), 142–159. https://doi.org/10.55982/openpraxis.16.2.649

Mediana Jr., N. L., Funa, A. A., & Dio, R. V. (2025). Effectiveness of inquiry-based learning (IbL) on improving students' conceptual understanding in science and mathematics: A meta-analysis. *International Journal of Education in Mathematics, Science and Technology, 13*(2), 532–552. https://doi.org/10.46328/ijemst.4769

Meyer, J. G., Urbanowicz, R. J., Martin, P. C., O'Connor, K., Li, R., Peng, P. C., ... & Moore, J. H. (2023). ChatGPT and large language models in academia: opportunities and challenges. *BioData mining*, *16*(1), 20.

Mo, Z., & Crosthwaite, P. (2025). Exploring the affordances of generative AI large language models for stance and engagement in academic writing. *Journal of English for Academic Purposes, 75*, Article 101499.

https://doi.org/10.1016/j.jeap.2025.101499

Moorhouse, B. L., Wan, Y., Wu, C., Wu, M., & Ho, T. Y. (2025). Generative AI tools and empowerment in L2 academic writing. *System, 133*, Article 103779. https://doi.org/10.1016/j.system.2025.103779

Nelson, A. S., Santamaría, P. V., Javens, J. S., & Ricaurte, M. (2025). Students' perceptions of generative artificial intelligence (GenAI) use in academic writing in English as a foreign language. *Education Sciences, 15*(5), 611. https://doi.org/10.3390/educsci15050611

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human–AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education, 49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

OpenAI. (2022, November 30). *Introducing ChatGPT.* https://openai.com/blog/chatgpt

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis, 18*(10), 2119–2126. https://doi.org/10.11124/JBIES-20-00167

Ramallosa, M., Funa, A. A., Geron, C. A., Ibardaloza, R. P., & Prudente, M. S. (2022). Meta-analysis on the effectiveness of argument-based learning on students' conceptual understanding. In *Proceedings of the 2022 13th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E '22)* (pp. 128–133). ACM. https://doi.org/10.1145/3502485.3502495

Rodafinos, A. (2025). The integration of generative AI tools in academic writing: Implications for student research. *Social Education Research, 6*(2). https://doi.org/10.37256/ser.6220256517

Stanford, J. (2025, March 17). How disruptive is AI in students' academic writing in higher education? *Studies in Technology Enhanced Learning, 4*(1). https://doi.org/10.21428/8c225f6e.0390b853

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

van Niekerk, J., Delport, P. M. J., & Sutherland, I. (2025). Addressing the use of generative AI in academic writing. *Computers and Education: Artificial Intelligence, 8*, Article 100342. https://doi.org/10.1016/j.caeai.2024.100342

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Wang, L., & Ren, B. (2024). Enhancing academic writing in a linguistics course with generative AI: An empirical study in a higher education institution in Hong Kong. *Education Sciences, 14*, 1329. https://doi.org/10.3390/educsci14121329

Watermeyer, R., Phipps, L., Lanclos, D., & Knight, C. (2024). Generative AI and the automating of academia. *Postdigital Science and Education, 6*, 446–466. https://doi.org/10.1007/s42438-023-00440-6

Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education, 21*, Article 21. https://doi.org/10.1186/s41239-024-00453-6

Zamorano, C. (2025). Enhancing academic writing in English language education through generative AI integration. *Research Studies in English Language Teaching and Learning, 3*(3), 424–447. https://doi.org/10.62583/rseltl.v3i3.87

Zhang, J. S., Yoon, C., Williams, D. K. A., & Pinkas, A. (2024). Exploring the usage of ChatGPT among medical students in the United States. *Journal of Medical Education and Curricular Development, 11,* 1–7. https://doi.org/10.1177/23821205241264695

Zhao, X., Cox, A., & Chen, X. (2025). The use of generative AI by students with disabilities in higher education. *The Internet and Higher Education, 66*, Article 101014. https://doi.org/10.1016/j.iheduc.2025.101014

Zizka, L. (2025). "It looks good enough": Recognizing the quality of generative AI output in academic writing tasks in higher education. *Journal of Hospitality & Tourism Education*. Advance online publication. https://doi.org/10.1080/10963758.2025.2496663

Appendix A. Summary of Studies on GenAI for Academic Writing in Higher Education: Contexts, Methods, Tools, Benefits, Challenges, and Gaps (2024–2025; n = 25)

#	Publication details		Country / Region	Population	Study context	Study Design / Methodology	GenAI tool/s used	Specific applications in academic writing	Reported benefits and opportunities	Identified challenges and ethical issues	Stated research gaps or recommendations for future studies
	Author	**Year**									
1	Cordero et al.	2025	Ecuador	University teachers (n = 184).	Series of workshops and courses for university instructors integrating GenAI tools; data gathered during these trainings; Canvas Learning Management System (LMS) used	Empirical mixed methods: surveys, observations, discussion forums, practical activities; descriptive statistics and thematic content analysis; exploratory analysis.	ChatGPT (Chat Generative Pre-trained Transformer), Gemini, Claude (text-generation tools).	Writing improvement; text summarization; feedback/idea generation; support for academic research and thesis preparation.	70% reported improved writing; 62% resource generation; 48% more efficient lesson planning; 45% summarization; 38% feedback/ideas/code; increased teacher confidence after workshops (+30%).	Ethical concerns and integrity issues (veracity 17%, privacy 30%, misinformation 41%, intellectual property 12%); only 19% confident identifying AI-generated text; need for AI-text detection though current tools are "not yet reliable."	Gap: practical and responsible classroom implementation guidance in higher ed. Recommendations: ongoing training, clear institutional policies, constant evaluation, and responsible use; strategies to distinguish human vs AI text.
2	Han	2025	China	300 undergraduates (65% humanities, 35% STEM) and 45 educators across 8 universities.	18-week academic-writing study across eight universities.	Mixed-methods, longitudinal: Phase 1 baseline (no AI); Phase 2 with three groups—A: full AI, B: post-draft technical checks only, Control: no AI; Phase 3 interviews/focus groups. Measures: Writing Competency Index (argumentation 0–50; originality 0–	ChatGPT; Grammarly.	Grammar correction, citation formatting, idea generation; Group B limited to post-draft technical checks; some students used AI-generated structures.	Grammar correction, citation formatting, idea generation; Group B limited to post-draft technical checks; some students used AI-generated structures.	Lower originality with broad AI use (e.g., Group A 21/30 vs Control 28/30); dependency (73% of Group A skipped pre-writing); 58% couldn't distinguish AI vs human drafts; humanities students reported loss of narrative voice (27%).	Gaps: need longitudinal cognitive data; ethical frameworks for AI co-authorship; address disciplinary disparities; cross-cultural work. Recommendations: phased AI use, transparency/disclosure protocols, AI-literacy modules, dynamic assessment emphasizing

						30; technical accuracy 0–20) and AI Dependency Metric; thematic analysis (NVivo) of 150 interviews.					idea originality, and equity interventions (e.g., subsidizing/ open-source software [OSS] tools).
3	Hysaj et al.	2025	United Arab Emirates	24 multicultural undergraduate students; English not a native language	Academic writing class at a United Arab Emirates (UAE) institution; data collected Apr–Aug 2023; university had academic-integrity policy but no formal GenAI policy yet (students had access to free platforms).	Qualitative; four focus groups; reflexive thematic analysis (Braun & Clarke).	Not specified by brand; participants referred generically to “GenAI” and “translator” tools in their writing processes.	Simplifying/judging journal articles; translation and paraphrasing; learning grammar/mechanics; improving vocabulary; completing written assessments.	Tools seen as practical, user-friendly, accessible, and expedient; boosted confidence, independence, and self-efficacy.	Dependency on GenAI; false information and grammatical errors; lack of reliability; disconnect between anti-plagiarism attitudes and unethical GenAI use; subjective marking concerns; limited institutional guidance (no formal GenAI policy during study).	Calls for further research on paraphrasing tools in students’ academic writing; recommends clear GenAI guidelines and PD for educators, student training/resources (especially for multicultural students), and reconsidering grammar in marking with more feedback on evaluative judgement and critical thinking.
4	Jensen & Jensen	2025	Denmark	Six Students in higher education	Demonstrates how students use GenAI and LLM for essay and assessment writing, and how academic writing instructors can respond	Teaching practice paper with case studies (three sets: chatbot essays, AI agent with feedback, and local LLMs)	ChatGPT (GPT-4), GPT-4 Application Programming Interface (API)(gpt-4-1106-preview), Sakura Solar Instruct (local LLaMa-based model)	Generating essays with prompt engineering, simulating writing tutor feedback cycles, running local models for essay writing	Enhancing efficiency and text quality, supporting multilingual writing, simulating tutor review, providing structured essay content, customizing outputs	Academic integrity concerns (misuse, plagiarism, contract cheating), AI detection tool limitations, loss of foundational skills, risk of false information, equity issues in technical skills	Instructors need to understand AI capabilities, develop policies, create alternative assessments beyond traditional essays, ensure equitable AI literacy among students
5	Johnston et al.	2025	United Kingdom	30 university students; 66.7 percent	Task focused on the planning & researching stage of	Task-based observational study with screen capture; reviewers	ChatGPT, Perplexity, Copilot, QuillBot, Grammarly,	Most students used GenAI to ask for explanations or definitions	Perceived advantages included speed, idea generation/struct	Issues observed: plagiarism via paraphrasing, use of unverifiable	Raise awareness of tools beyond ChatGPT; teach

				undergraduate and 33.3 percent postgraduate taught; mixed disciplines	essay-style assignments; 50-minute, screen-recorded session in a library study room; student chose 1 of 3 subject-appropriate essay questions and narrated their process.	produced step-by-step tables; some double-reviewed for consistency; non-parametric stats (Mann–Whitney U, Spearman's ρ); ethics approval U. Liverpool 12737.	PopAI, JotBot, Vertex	and to generate an essay plan or structure	ure help, interactive refinement; Copilot/Perplexity valued for showing sources.	sources, one participant used only GenAI; students did not cite GenAI in plans; concern that ChatGPT references can be wrong.	limitations and fact-checking; allow students to explore GenAI; include Google and Google Scholar in teaching; note students' use of MyBib; offer self-analysis or quiz via the Joint Information Systems Committee (Jisc) Discovery Tool; future research to focus on specific student groups, study prompt literacy, and track effects of AI integration into search engines and office software
6	Karahan Adalı & Bilgili	2025	Türkiye	226 university students from 16 associate and undergraduate departments (55.2% from informatics; age 18–27+, 56% female, 44% male)	Perceptions, attitudes, purposes, and ethical concerns of university students in Türkiye toward GenAI tools in higher education	Quantitative; descriptive survey model; cross-sectional; purposive (maximum diversity) sampling; statistical analyses included factor, correlation, and cluster analysis	ChatGPT, Claude, Bard (and other LLMs mentioned)	Writing essays, generating study materials, summarizing texts, improving grammar, brainstorming, problem solving, translation, coding assistance, data analysis, language learning, creative writing	Enhanced academic productivity; personalized learning; improved language and writing skills; engaging and interactive learning; inclusive learning environments	Misinformation; plagiarism risk; AI-enabled inequalities; academic dishonesty; bias; data privacy concerns; gender and discipline-based ethical perception differences	Develop discipline-specific AI guidelines; address ethical concerns while promoting responsible use; consider demographic differences; provide targeted AI literacy and skill development; future research on longitudinal changes in attitudes and effectiveness of integration strategies
7	Kim, Lee et	2025	China	36 Chinese Master of	One-on-one think-aloud	Purposeful sampling: AI	Custom Unity-based GenAI	Used during ideation/drafti	High-AI-literacy group showed a	Challenges noted include	Recommend examining

	al.			Science (MSc) Digital Education students; 3 males and 33 females; age 21–30.	academic essay task in a university lab; training on GenAI prompts and think-aloud; assigned one of two topics; most finished within ~1.5 hours.	literacy questionnaire split at the median score of 5.09 into high- and low-literacy groups; data sources included think-aloud transcripts, screen recordings, and chat logs; coded student and AI interaction (SAI); statistical analyses included the chi-square test, Epistemic Network Analysis (ENA), and the Mann-Whitney U test; expert rubric scoring	writing system using GPT-4 (gpt-4-0613, OpenAI).	ng and to invite, adopt, and refine AI suggestions; planning and evaluation integrated into the writing workflow.	collaborative SAI pattern and significantly higher scores for content, structure, and expression; higher total writing score.	lower-literacy students interacting less with AI and expressing negative emotions; barriers such as weak prompting/critical evaluation and potential over-reliance discussed in the literature; ethical note: IRB approval obtained.	broader learner traits (e.g., baseline academic skills) and using multimodal data (eye-tracking, electroencephalography [EEG]); suggests expanding beyond a single course.
8	Kim, Yu et al.	2025	China	n = 20 higher-education students (7 bachelor's, 8 master's, 5 doctoral); 10 female and 10 male; International English Language Testing System (IELTS) scores 5 to 7.5; split into high and low AI literacy groups	English-medium instruction (EMI) Sino-British university; students completed an IELTS-type Academic Writing Task 2 before interviews (250+ words, timed).	Semi-structured interviews conducted via Zoom lasting 60 to 90 minutes after a writing task; purposeful plus snowball sampling; mixed inductive and deductive thematic analysis with member checking	ChatGPT-4 embedded "Writing With GPT (WWG)" system (Unity); hidden tutor prompt + open student prompt; 60-min writing, 20-min system familiarization.	Used across the writing process: ideation, planning/outline, drafting, and revision.	Improved writing quality, speed/efficiency, and topic/content knowledge; affective gains: enjoyment, question-generation, perceived support, self-efficacy.	AI-related barriers: hallucinations; lack of contextual understanding, higher-order thinking, human awareness, cultural awareness, relationship skills, pedagogical skills; interoperability gaps; lack of explainability. Student-related: low AI-literacy, negative attitudes, limited higher-order thinking, weak topic knowledge,	Implications/recs: build human-centered, explainable GenAI for learning; embed AI-literacy & prompt-engineering across disciplines; design to foster higher-order thinking. Future research: larger, more diverse samples; additional tasks and settings; longitudinal classroom

										weak writing skills. Task-related: time constraints.	studies.
9	Kofinas et al.	2025	United Kingdom	Eight higher education academics (markers) from two UK universities; undergraduate business students' assessment outputs	UK higher education, business school context; impact of GenAI on academic integrity and detectability in authentic assessments	Mixed-methods, two-phased within-subjects experimental design; quantitative marking comparison + qualitative semi-structured interviews; thematic analysis	ChatGPT 3.5	Modifying existing student assessments; generating entire assessments based on briefs; simulating student use of GenAI to complete assignments	Demonstrated that GenAI can produce authentic-looking assessments that pass academic scrutiny; some grade improvements in lower bands; potential to enhance efficiency in assessment completion	Low detectability of GenAI use; false positives and negatives in detection; risk of undermining academic integrity; bias in detection tools; ethical concerns (bias, hallucinations, over-reliance)	Authentic assessments alone insufficient to safeguard integrity; need for rethinking assessment design; shift toward process-based and synchronous assessments; policy focus on design rather than format
10	Mo & Crosthwaite	2025	United Kingdom	30 British tertiary students (Levels 3 and 4; final-year undergraduates & master's students).	Comparison of stance and engagement features in human vs. AI-produced academic writing across six disciplines (Classics, Archaeology, History, English, Philosophy, Linguistics)	Corpus-based comparative study; annotation of stance and engagement features (Hyland, 2005) using UAM Corpus Tool (Universidad Autónoma de Madrid Corpus Tool, linguistic annotation software); statistical analysis (ANOVA, Kruskal-Wallis, Mann-Whitney U, LSD tests); discipline-specific and model-specific comparisons	ChatGPT 4.0, ERNIE Bot 4.0 (Enhanced Representation through Knowledge Integration), Meta AI Large Language Model Meta AI (LLaMA) 3.1	Generation of academic essays based on British Academic Written English (BAWE) essay prompts and contextual notes to replicate human academic writing in terms of stance and engagement	Some disciplinary alignment of stance features between AI and human writing; potential for AI to produce human-like discourse in certain contexts; useful for English for Academic Purposes (EAP) pedagogy to illustrate differences in stance/engagement	Narrower, more repetitive stance and engagement repertoire in AI writing; over-reliance on certain markers (e.g., "significant"); underuse of boosters and hedges; weaker engagement, especially in philosophy essays; limitations in discipline-specific rhetorical conventions	Need for further research on AI stance/engagement across more disciplines; refining LLM training to improve variety and context sensitivity; implications for EAP instruction and automated AI-text detection
11	Moorhouse et al.	2025	Hong Kong	n = 21 postgraduate second-language writers from Mainland China; 5	One-year MEd, English-medium instruction; first month of studies; course-level	Qualitative study using semi-structured focus-group interviews and individual interviews;	ChatGPT; other LLMs (incl. a popular China-based LLM referenced by participants)	Linguistic support; explaining difficult concepts; summarizing literature; polishing/corr	Students perceive GenAI as empowering/assistive; improves efficiency, organization, confidence;	Accuracy issues and hallucinated references; risk of over-reliance/loss of critical thinking; fairness	Embed critical digital literacy in postgraduate writing instruction; design assessment policies that

				male and 16 female; first language Mandarin; International English Language Testing System overall proficiency noted (score not specified)	GenAI policies with AI-use declarations possible	preference-selection task; interviews conducted in Mandarin; 7 focus groups with 2 to 3 participants each plus 3 individual interviews; duration 21 to 65 minutes; descriptive statistics for the preference task; inductive thematic analysis; ethical approval obtained		ecting language; generating structure and subheadings; brainstorming/idea generation	prefer flexible “full AI with declaration” policies for accountability/fairness; training on how to use AI seen as useful/needed	concerns with bans and unreliable AI detection; questions about burden/what to declare	give students agency with transparency; replicate in other contexts beyond Hong Kong high-resource setting
12	Nelson et al.	2025	Ecuador	56 undergraduate B1 English as a Foreign Language (EFL) students; two sections of a required B1 EFL course	Partial English-medium instruction (EMI) science, technology, engineering and mathematics (STEM)–only university; classroom-dependent English proficiency; academic writing in English.	Anonymous multiple-choice survey (Microsoft Forms), 11 questions on AI use in writing; administered in class with consent.	Focus on ChatGPT (students’ perceptions of ChatGPT for creating/improving second-language [L2] writing).	Students reported ChatGPT as valuable, saves time, helps with writer’s block, supports brainstorming; also used as a language assistant for synonyms/clarifying meaning/formal phrasing of ideas they already intend to express.	71% saw ChatGPT as valuable (saves time; combats writer’s block; brainstorming). Students want institutional support/awareness and digital-literacy guidance on ethical use.	Concern about hindering writing development and producing pseudo-success; dependency risk; perceived increase in academic dishonesty. Many believed chatbot-written work is easily detectable. Submitting chatbot-generated text was viewed as academic dishonesty, while machine-translated Spanish to English was less often viewed as dishonest.	The study highlights an under-represented population (rural STEM EFL context in South America) and calls for policies, teaching practices and AI-literacy adapted to this setting. It suggests teacher training and student support, and a methodological refinement to reduce power-dynamic bias by using a third-party facilitator or surveying outside class hours in future research.
13	Rodafinos	2025	Greece	n = 45 first-year undergradu	Students completed literature-	Position paper with a qualitative	Various LLMs (e.g., ChatGPT,	Literature search and summarizing;	Time efficiency; improved scientific	Need to verify AI outputs; inaccuracies	Calls for more empirical research on

				ates in a Research Methods course	review papers; AI use permitted; each submitted a 1–2-page reflection on AI use.	case study; reflections analyzed via thematic analysis.	Gemini, Copilot, Perplexity, DeepSeek, Claude); AI translation tools; AI used alongside PubMed & Google Scholar for search.	highlighting methods; suggesting sources; drafting introductions, literature reviews, and conclusions; idea formulation; organization and coherence; grammar and style; American Psychological Association (APA) formatting; structuring findings and references; translation English ↔ Greek; critiquing titles according to APA style; aiding statistical interpretation; refining research questions	accuracy/clarity/cohesion/presentation; supportive “assistant” role; ease of use. (Also notes institutional benefits from AI adoption.)	and vagueness; reliance on non-academic sources; grammar and syntax issues AI should not replace critical thinking; transparency and citation of AI use urged; risk of fabricated references and citations (hallucinations)	long-term impacts and tool evaluation (case studies, controlled experiments, qualitative studies); proposes activity theory as a framework; institutional recommendations: clear policies, AI literacy in curricula, faculty training, and assessment redesign to demand critical thinking.
14	Stanford	2025	United Kingdom	Six master’s students at a university in the Midlands.	Higher education academic writing in social sciences	Qualitative case study using semi-structured interviews; guided by Constructivist Theory and Unified Theory of Acceptance and Use of Technology (UTAUT) framework; thematic analysis	ChatGPT, Grammarly, Google Scholar, Google Translate	Grammar/style correction, brainstorming, planning, overcoming language barriers, literature search, citation management	Improved efficiency; support for non-native speakers; enhanced fluency and confidence; idea generation; overcoming writer’s block; organizational support; personalized feedback	Over-reliance on AI; lack of originality; ethical concerns (authorship, plagiarism); limited digital literacy; unclear institutional guidelines; limited access to premium tools; data security; potential misinformation; systemic bias	Expand sample to include wider disciplines and settings; integrate AI education in curriculum; strengthen policy communication; support non-native speakers; promote critical engagement with AI; conduct longitudinal

											studies on AI adoption impacts
15	van Niekerk et al.	2025	Norway	Undergraduate computing students in an academic writing skills course; 173 enrolled (41% female). Pre-survey N=159; post-survey N=95 (voluntary, anonymous).	First-semester academic writing skills course spanning Cyber Security, Digital Forensics, and Applied Data Science.	Active-learning, multi-stage intervention grounded in the Technology Acceptance Model (TAM). Students generated and then critiqued an AI-written essay. Two Likert-scale questionnaires (pre and post) analyzed with the McNemar–Bowker test; instrument reliability reported using Cronbach's alpha.	ChatGPT (GPT-3.5), the most advanced freely accessible model during Aug–Sep 2023 data collection.	Students prompted ChatGPT to produce a 1,000-word essay with ≥5 in-text citations, then evaluated it against seven academic-writing characteristics (e.g., structure/flow, evidence & references, formal tone).	Active-learning intervention significantly reduced over-reliance and helped students identify appropriate uses of ChatGPT; the paper positions ChatGPT as a tool for polishing (e.g., improving form and tone) rather than producing academic text.	Hallucinated/incorrect references requiring verification; limitations in critical insight; possibility of biased/incorrect outputs; risk that over-reliance undermines students' ability to develop original ideas.	Replicate with newer GPT versions; examine durability of effects via longitudinal follow-ups; consider other subjects (e.g., programming).
16	Zamorano	2025	Chile	40 first-year English for Academic Purposes (EAP) undergraduates; Spanish-speaking English as a Foreign Language (EFL) learners at B1 to B2 on the Common European Framework of Reference for Languages (CEFR); randomized into an AI group (n = 20) and a	Six-week instructional period at a mid-sized private university in Santiago, part of a first-year EAP writing course	Convergent parallel mixed methods; quasi-experimental pre and post design with repeated-measures ANOVA, reporting partial eta squared (ηp^2) effect sizes; qualitative reflections thematically analyzed in NVivo qualitative analysis software; inter-rater reliability Cohen's kappa (κ) = .84; Institutional Review Board (IRB)	ChatGPT (as drafting/revising aid during writing workshops)	Guided prompting for ideation, outlining/structuring, lexical choice/expansion, and editing; tasks: expository, argumentative, cause–effect essays; weekly reflection journals	Significant gains vs control in vocabulary use, structural organization, and audience awareness; increased motivation and enjoyment; support for idea generation, outlining, and vocabulary discovery	Over-reliance on AI; risks of passive learning and reduced analytical thinking; explicit guidance to use AI to support—not replace—student ideas; consent, anonymity, secure data storage; disclosure and academic integrity discussed	Implement AI with guided execution, reflective practice, and ethics-focused instructional design; model effective prompting; treat GenAI as scaffolding/drafting aid rather than text generator; explore longer-term effects and transferability to other contexts (implied)

				control group (n = 20)		approved					
17	Zhao et al.	2025	United Kingdom	124 students with disabilities from a UK university	Investigating GenAI use in academic writing among students with disabilities, including perceived benefits, barriers, and recommendations for support	Online survey distributed to all students with disabilities; 124 valid responses; quantitative data analyzed descriptively; qualitative data analyzed using inductive content analysis	Various GenAI tools (not specified by brand)	Used to address barriers related to disabilities in academic writing, improve productivity, assist with writing challenges	Helped overcome certain disability-related barriers; supported academic writing tasks	Risk of over-reliance; issues of accuracy; concerns about cost and accessibility	Recommend enhancing university support for students with disabilities in AI use; further research on disability-specific AI applications in education
18	Zizka	2025	Switzerland	Graduate students; final usable outputs n = 88 (initial 90; two didn't run the second prompt).	Graduate AW class; academic-integrity session where students queried ChatGPT, then requested a "graduate level" rewrite; outputs sent to professor for analysis.	Classroom exercise with descriptive text analytics: Microsoft Word statistics and readability measures; corpus word-frequency analysis and word clouds; detailed content analysis of three random samples	ChatGPT-3.5 (free version).	Students asked ChatGPT to write a 500-word argumentative response with five in-text citations and references in American Psychological Association 7th edition (APA 7) style, then to rewrite the output "at a graduate level"	Opportunities discussed include integrating GenAI within redesigned Academic Writing courses and focusing on PACE: Prompt engineering, Authenticity and Accuracy, Critical thinking, and Evaluation and Assessment	Hallucinations/faulty reasoning/poor referencing noted; minimal differences between "standard" and "graduate" outputs; identical structures risking plagiarism suspicions; frequent incorrect or fabricated references; students unable to validate quality.	Replicate with other programs/levels; test other GenAI tools; deeper critical analysis of outputs. Also, curricular recommendations (innovative AW courses; incorporate GenAI with evaluation; or bans and their limits).
19	Campbell & Cox	2024	United States	Graduate learners in an education department: N=25 (23 survey; 25 in discussion); 17 women/6 men; all employed full-time; 5 joined a 24-week	Perceptions and use of GenAI chatbots in higher education to inform andragogical practice; only 21.7% reported using chatbots for academic purposes	Mixed methods (descriptive plus qualitative): online questionnaire using Google Forms and immediate class discussion; exploratory content analysis with Linguistic	ChatGPT, Google Gemini, Microsoft Copilot AI, Claude	Used to format references, draft responses to prompts (some asked ChatGPT to "write" from their ideas), make presentations, find & summarize information; some courses required	Timesaving; boosts writing self-confidence/efficacy; increases capacity; helps look up answers and research topics	Main concerns: plagiarism and cheating; accuracy/quality of AI outputs; many learners more negative than positive on ethics items	Future work: more qualitative studies on learner experiences; embed Gen-AI in curriculum and evaluate outcomes (with control groups); replicate across disciplines; guidance for

				follow-up discussion		Inquiry and Word Count, version 22 (LIWC-22), plus human coding; quasi-statistics; convenience sample; 24-week follow-up small-group discussion		deconstructing AI outputs for accuracy			educators to set clear policies and teach critical evaluation/source-checking
20	Johnston et al.	2024	United Kingdom	2,555 university students (≈8.86% of student body)	Gather student perspectives on GenAI to inform updates to the Academic Integrity code of practice	Three student-led focus groups to vet items, then a 9-question online survey (March–April 2023); ethics approved; analysis in Excel/SPSS with Mann–Whitney U and Goodman & Kruskal's gamma	Grammarly, ChatGPT (highest awareness: Grammarly 88.5%, ChatGPT 68.9%)	Most-suggested uses: grammar/spelling/punctuation; understanding concepts; planning; summarising text; referencing; also "academic" use category included grammar checking and reference formatting	54.1% supportive/somewhat supportive of using tools like Grammarly; students want clear guidance; study argues against blanket bans and suggests co-creating guidance with students	Strong opposition to using ChatGPT to write entire essays (majority unsupportive); tool limitations and misinformation noted (e.g., referencing errors, bias); fairness concerns raised	Preferred response: a university-wide policy (41.1%); educate students; ensure equitable access (consider disabled and international students); do not ban; future research to examine links between lower writing confidence and GenAI use
21	Maphoto et al.	2024	South Africa	Module enrolment ≈14,000; sample: 12 lecturers (of 20), 12 students (of 20), 10 markers (of 30)	WRI124 Academic Writing module in an online distance-learning (ODeL) setting; Moodle LMS & MS Teams; early ChatGPT uptake in 2023 assessments	Qualitative, exploratory/interpretive; phenomenological design; triangulation: email interviews (lecturers), FGD (students), WhatsApp group (markers); purposive sampling; data collected Nov–Dec 2023	ChatGPT (observed heavy student use in online assessments); broader framing refers to LLMs/AI writing tools	Draft correction/comparison, grammar/structure support, brainstorming/idea generation, feedback and editing (as discussed by stakeholders and literature)	Positive impact on teaching/learning; innovative opportunities; supports error identification/structural improvements; potential motivational tool	Risks of academic misconduct/over-reliance; authenticity/privacy concerns; plagiarism; digital divide among staff vs students; need for responsible use	Recommend balanced integration with clear guidelines, training, and user-friendly interfaces; transparent communication on capabilities/limits; explore AI as motivational tool; future research across more modules/disciplines and different AI systems
22	Nguyen et al.	2024	Finland and New	n = 10 doctoral	Online writing task	Online experimental	ChatGPT (GenAI-	Students prompted for	Iterative, highly interactive use	Linear copy-paste with little	Calls for deeper study

			Zealand	students in English-medium doctoral programs in Information Systems (IS) or Learning and Educational Technology (LET)	via Zoom: 500-word essay on AI in education; 30 minutes; could use any tools (ChatGPT, Google Scholar); screen-recorded interactions.	design using AI-driven learning analytics with three layers: (1) quantitative content analysis, (2) sequence analysis using a Hidden Markov Model (HMM) with hierarchical clustering, and (3) process mining	powered assisting tool).	content, outlines, grammar correction, feedback, and follow-ups (e.g., PromptContent, PromptOutline, PromptCorrection, PromptFeedback, PromptFollowUp); they read, copied, and pasted AI-generated text into essays.	(“Structured Adaptivity”) associated with higher performance (Type 1 > Type 2; t=2.4011, p≤0.05); effective patterns include multitasking (prompting while article-searching) and reflective integration/editing of generated text.	critical assessment (“Unstructured Streamline”) linked to lower performance; some deletion of previously pasted AI text. Ethical issues flagged for future work include data privacy and potential AI bias.	of human-AI collaboration (cognitive & metacognitive factors), use of eye-tracking or verbal reports, development of more appropriate assessment tools for learning outcomes, and further work on privacy and AI bias.
23	Wang & Ren	2024	Hong Kong	140 undergraduates in an “Introduction to Linguistics” course; survey subset n = 98; 29 Wikibook chapters analyzed.	13-week collaborative multimedia Wikibook project integrating AI within course writing tasks.	Mixed-methods: analysis of 29 chapters and students’ AI interaction logs; online survey (5-point Likert) with 98 responses.	ChatGPT-3.5; Padlet AI image generation.	Identified roles: supporting opinions (definitions, arguments, concept analysis), linguistic repository, organizational ideas; used most for drafting/revising content.	Students reported usefulness for idea generation, vocabulary, grammar, and organization, with perceived productivity gains; the teacher observed improved content richness, visuals, and grammatical accuracy compared with the prior year without AI	Risks noted include plagiarism or misuse, over-reliance, potential misinformation, and the need for ethical and transparent integration guidelines	Literature gap on combined use of ChatGPT and AI image generators in academic writing; future studies should include more disciplines and education systems
24	Watermeyer et al.	2024	United Kingdom	University staff in the UK; analysis focuses on academics (n = 284 of n = 428 total respondents)	Use of GenAI by United Kingdom academics amid audit culture, overwork, and post–coronavirus disease 2019 (COVID-19) higher-education conditions	Anonymous online survey (Qualtrics), launched June 2023 for 2 months; convenience sampling via mailing lists/social media; descriptive stats + thematic analysis of open responses	GenAI / LLMs (e.g., ChatGPT)	Speeding up the writing process; generating reflective pieces (for example, Postgraduate Certificate [PGCert] reflections); summarizing reports; copywriting; adding disclaimers regarding AI	Labor-saving on administrative/menial text tasks; productivity acceleration; potential “clearing space” for more meaningful work; possible status equalizer; reclaiming some autonomy	Threats to authorship/integrity and quality; expectation/pressure to produce more (work intensification); risk of ethics being overlooked; potential to exacerbate inequalities/stratification; fragmentation of collegial work;	Authors note this is a “snapshot” and a platform for future work; suggest further research across roles, disciplines, services, institutions, and sectors; separate analysis of professional-services staff

								assistance		churn/hyper-productivity concerns; calls for AI-use disclaimers	
25	Yusuf et al.	2024	Global, multicultural survey across 76 countries in higher education.	Students and lecturers; n = 1,217 after data cleaning.	Higher education; examines usage, benefits, concerns of GenAI from a multicultural perspective	Embedded mixed-methods, quantitative-dominant online survey using convenience sampling; descriptive and inferential analyses conducted in Statistical Package for the Social Sciences (SPSS); qualitative content analysis; cultural linkage via Hofstede’s cultural dimensions (for example, ordinal logistic regression and network-graph analyses)	Familiarity measured for: ChatGPT; GrammarlyGo; Bard; DALL·E; also JukeBox, Synthesia, Stable Diffusion, Midjourney, ChatSonic, YouChat	Information retrieval; paraphrasing; literature search & summarized reading; brainstorming/starting points; writing support & generation (codes, essays, poems, scripts)	Personalized & immediate learning support; strong query response; brainstorming aid; literature search & summaries; writing support; promotes equity & access	Factually inaccurate or out-of-context outputs; bias/unfairness; heavy reliance on online sources; limited emotional intelligence; potential for cheating & plagiarism; over-reliance hindering growth; threats to academic integrity	Gap: lack of multicultural perspectives in HE. Recommendations: responsible use with robust, culturally responsive policies; avoid one-size-fits-all; continue research & dialogue; improve sampling balance and consider personalized survey links in future work