

19. Sedykh T.V., Korshunova V.V., Sosnovskaia A.A., Grigorovech P.N., Bugaeva A.A. *Practice-Oriented Approach to Development of Leadership Competencies of Honors Students: The Project «The Territory of Intellectual and Liberal Inventions»* // *Journal of Siberian Federal University. Humanities & Social Sciences*. 2021. DOI: 10.17516/1997-1370-0823.
20. Hanbury G.L. *Advancing University Core Values by Developing an Honors College* // *The Journal of the National Collegiate Honors Council*. 2015. T. 16. C. 91-97.
21. Regan, P. M., Jesse, J. *Ethical challenges of edtech, big data and personalized learning: twenty-first century student sorting and tracking* // *Ethics and Information Technology*. 2018. DOI: 10.1007/s10676-018-9492-2.
22. Deng Y., Lu D., Chung C.-J., Huang D., Zeng Z. *Personalized Learning in a Virtual Hands-on Lab Platform for Computer Science Education* // *IEEE Frontiers in Education Conference (FIE)*. 2018. DOI: 10.1109/FIE.2018.8659291.
23. Shetty S., Shetty A., Salian A.B., Umesh P., Gangadharan, K. *Experiential Learning of Strength of Materials and Fluid Mechanics using Virtual Labs* // *IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. 2020. DOI: 10.1109/CONECCT50063.2020.9198678.
24. Cherner Y., Cima M., Barone P., Van Dyke B.R., Lotring A.O. *Interactive and Adaptable Cloud-based Virtual Equipment and Laboratories for 21st Century Science and Engineering Education* // *Proceedings of the 3rd International Conference on e-Learning and Virtual Science*. 2020. DOI: 10.29007/7wf8.
25. Osipovskaya E. *Using Google Trends to analyze top EdTech-trends 2020 in Russia and worldwide* // *RUDN Journal of Informatization in Education*. 2021. DOI: 10.22363/2312-8631-2021-18-4-291-304.
26. Rajini S., Hari Prasad B., Upendrasingh A. *Alumni Management and Networking System* // *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*. 2023. DOI: 10.1109/ICAECA56562.2023.10200060.

IRSTI 14.35.09

<https://doi.org/10.51889/2959-5762.2025.87.3.003>

Jandildinov M.,<sup>1\*</sup>  Yersultanova G.<sup>1</sup> 

<sup>1</sup>*Abai Kazakh National Pedagogical University, Almaty, Kazakhstan*

## AI-ASSISTED SCHOLARLY WRITING IN EDUCATION: A SCOPING REVIEW (2019–2024)

### *Abstract*

The past five years have seen a rapid rise in the use of artificial intelligence (AI) to support scholarly writing in educational contexts. This scoping review examines literature from 2019–2024 on AI-assisted writing in education. We identify key concepts and synthesize research methods used to study this phenomenon. Prominent concepts include generative AI large-language models (e.g. ChatGPT) and automated writing evaluation (AWE) tools. The review followed PRISMA-ScR guidelines, searching major academic databases (e.g. Scopus, Web of Science) and applying predefined inclusion criteria. A flowchart (Figure 1) summarizes the study selection and categorization process. We summarize representative empirical and review studies in Table 1. Results show that many recent studies focus on LLM-based writing assistants and writing feedback tools. Methods range from qualitative case studies and surveys to quasi-experiments, with a few systematic reviews (e.g. Imran & Almusharraf, 2023) emerging. Common findings include improved writing efficiency and positive student perceptions, alongside concerns about plagiarism and the need for AI literacy. Gaps include a lack of longitudinal or large-scale controlled studies and inconsistent theoretical grounding. We recommend more rigorous designs, reporting of all outcomes (including null effects), and integration of learning theory. This review provides an integrated snapshot of current research on AI-assisted scholarly writing in education and suggests priorities for future work.

**Keywords:** AI-assisted writing, generative AI, automated writing evaluation, academic writing, educational technology, large language models.

М.К.Джандильдинов,<sup>1\*</sup>  Г.Т.Ерсултанова<sup>1</sup> 

<sup>1</sup>Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан

## БІЛІМ БЕРУ САЛАСЫНДА ЖАСАНДЫ ИНТЕЛЛЕКТ КӨМЕГІМЕН ҒЫЛЫМИ ЖАЗУ: ШОЛУ (2019–2024)

*Аңдатпа*

Соңғы бес жылда жасанды интеллект (ЖИ) оқу үдерісінде ғылыми жазуды қолдауға белсенді қолданылып келеді. Бұл шолу 2019–2024 жылдар аралығындағы білім беру саласындағы ғылыми жазуды ЖИ көмегімен іске асыруға арналған зерттеулерді қарастырады. Мұнда негізгі ұғымдар анықталды (мысалы, генеративті тіл модельдері мен автоматтандырылған жазылым бағалау жүйелері), зерттеу әдістері талданды. Негізгі ғылыми дерекқорлар бойынша іздеу жүргізілді және таңдалған мақалалар кестеленді. Нәтижелер бойынша ChatGPT сияқты модельдер мен жазылымға кері байланысты қамтамасыз ететін құралдар басым. Зерттеулер әдістері сапалық жағдайлық зерттеулер мен сауалнамалардан бастап квази-тәжірибелерге дейін өзгерді; сондай-ақ алғашқы жүйелі шолулар пайда болды. Зерттеулер жазу тиімділігінің артуын және студенттердің оң пікірін көрсетті, бірақ академиялық адалдық мәселелері туралы алаңдаушылық та айтылды. Зерттеуде уақытша және ауқымды бақылау кемшін тұстары анықталды. Болашақта зерттеулерді оқыту теорияларымен байланыстыра отырып, бар нәтижелерді, соның ішінде маңызсыз нәтижелерді де жариялау ұсынылады.

**Түйін сөздер:** жасанды интеллект, ғылыми жазылым, автоматтандырылған жазылымды бағалау, академиялық жазу, білім беру технологиялары, генеративті тіл модельдері.

Джандильдинов М.К.,<sup>1\*</sup>  Ерсултанова Г.Т.<sup>1</sup> 

<sup>1</sup>Казахский национальный педагогический университет имени Абая,  
г.Алматы, Казахстан

## НАПИСАНИЕ НАУЧНЫХ РАБОТ С ПОМОЩЬЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБРАЗОВАНИИ: ОБЗОР (2019–2024)

*Аннотация*

За последние пять лет стремительно возросло использование искусственного интеллекта (ИИ) для поддержки научного письма в образовательном контексте. В данном обзорном исследовании рассматривается литература 2019–2024 годов по теме написания текстов с помощью ИИ в образовании. Мы определяем ключевые понятия (например, генеративные языковые модели и системы автоматической оценки текста) и обобщаем методы исследований. Поиск осуществлен по основным базам данных, результаты отбора показаны на схеме (Рисунок 1). Представлены примеры эмпирических и обзорных работ (Таблица 1). Результаты свидетельствуют о том, что большинство исследований посвящено моделям типа ChatGPT и инструментам обратной связи по письму. Методы варьировались от качественных случаев и опросов до квазиэкспериментов; появились первые систематические обзоры. Обнаружены улучшения эффективности письма и положительная реакция студентов при использовании ИИ, наряду с опасениями за академическую честность. Выявлены пробелы (например, нехватка долгосрочных исследований и единых теоретических основ). Рекомендуются более строгое проектирование исследований и открытое представление всех результатов, включая незначимые.

**Ключевые слова:** ИИ-помощь при письме, научное письмо, автоматическая оценка письма, академическое письмо, образовательные технологии, генеративный ИИ.

**Introduction.** In recent years, artificial intelligence (AI) has become rapidly integrated into educational practices, including the writing process. The advent of powerful generative AI models like OpenAI’s ChatGPT (released in late 2022) has marked a potential paradigm shift: these models can produce coherent, human-like text and assist with drafting essays, research papers, and other academic writings. This surge in AI capability has generated intense debate in academia. Educators and researchers are exploring how AI tools might improve writing efficiency and learning outcomes, while also grappling with new ethical challenges such as plagiarism and academic integrity breaches. As one review notes, “the past five years have seen a rapid rise in the use of AI to support scholarly writing in educational contexts,” underscoring the timeliness of this topic.

Despite the growing interest, the scholarly understanding of AI-assisted writing in education is still at an early stage. Recent work has begun to map this terrain by examining two main categories of tools: large-language-model (LLM) writing assistants (like ChatGPT) and automated writing evaluation

(AWE) systems (such as ETS e-rater or Grammarly). For example, Imran and Almusharraf (2023) systematically reviewed 30 studies on ChatGPT’s role as a writing aid. Other studies have investigated traditional AWE tools’ effects on student writing performance. Empirical research to date is diverse: methods range from qualitative case studies and interviews to surveys and quasi-experiments, and the first systematic or narrative literature reviews are only now emerging. In general, however, much of the literature is fragmented. Few large-scale or longitudinal studies exist, and many papers lack a unified theoretical framework or consistent methodology.

*Research Aim.* This scoping review aims to provide a comprehensive overview of scholarship on AI-assisted scholarly writing in education from 2019 through 2024. Following PRISMA-ScR guidelines, we systematically collected peer-reviewed studies indexed in major databases (Scopus, Web of Science, etc.). We clarify key concepts (e.g., generative AI, AWE), catalog the research designs and contexts, and synthesize the principal findings and trends. By doing so, we seek to identify gaps and inconsistencies in the literature and to suggest best practices and future directions. In other words, our goal is to chart the current landscape of research on AI writing tools in education and to guide researchers and practitioners toward the most promising and rigorous lines of inquiry.

*Significance of the Study.* As generative AI continues to evolve, understanding its implications for writing pedagogy is crucial. This review offers an up-to-date synthesis of evidence, helping educators and technology specialists make informed decisions about integrating AI tools. For example, the findings highlight that while AI assistance can boost student motivation and efficiency in drafting, it also raises significant integrity issues and necessitates AI literacy training. By summarizing effective practices (such as structured human–AI collaboration and explicit guidance on tool use) and noting where research is weak, our work bridges the gap between AI research and educational practice. Ultimately, a clear picture of what is known—and not yet known—about AI-assisted writing will help institutions craft policies on academic integrity, design curricula that leverage AI’s benefits, and encourage future research that is methodologically sound and pedagogically grounded.

*Basic provisions.*

- *Focus on LLMs and Feedback Tools.* Recent studies have primarily examined large-language-model (LLM) based writing assistants (e.g. ChatGPT) and automated feedback tools (grammar/style checkers, writing evaluators). Researchers have used diverse methods – from qualitative case studies and surveys to quasi-experimental designs – and initial systematic reviews are appearing in this emerging field.

- *Positive Effects on Writing.* The bulk of evidence suggests that AI assistance can improve writing efficiency and quality. Many studies report that students perceive AI tools as helpful for idea generation, language editing, and overall productivity. For example, integrating AI feedback software into writing instruction has been linked to higher writing performance compared to traditional methods.

- *Academic Integrity Concerns.* Alongside benefits, widespread concerns about plagiarism and academic honesty were documented. Scholars consistently note that easy text generation by AI “raises concerns about increased risks of cheating” and undermines existing norms of authorship. This has led to calls for clear guidelines on acceptable AI use and training students in responsible use of these tools.

- *Research Gaps:* The review identifies important gaps in the literature. Very few studies are longitudinal or large-scale, and randomized controlled trials remain rare. Many papers lack explicit theoretical foundations. As a result, the evidence base is still fragmentary and may not generalize across different contexts (e.g. K–12 vs. higher education, first-language vs. English-language learners).

- *Recommendations for Future Research.* The authors emphasize the need for more rigorous and transparent research. This includes designing stronger studies (e.g. controlled experiments, mixed methods) and reporting all results (including non-significant findings). They also urge integrating learning theories to better interpret outcomes. By adhering to these best practices, future investigations can build a more robust understanding of how AI tools affect scholarly writing.

**Materials and Methods.** We conducted a scoping review following established guidelines (e.g. Tricco et al., 2018) to systematically gather and characterize studies on AI-assisted writing in education [1]. We searched academic databases (including Scopus, Web of Science, IEEE Xplore, and ERIC)

using combinations of keywords related to artificial intelligence (e.g. “AI”, “ChatGPT”, “generative writing model”, “automated writing evaluation”) and education or scholarly writing (e.g. “academic writing”, “education”, “student writing”). Inclusion criteria were: (1) peer-reviewed publications from 2019–2024; (2) empirical or conceptual studies focused on educational contexts of scholarly writing; and (3) use of AI-related tools or algorithms for writing assistance. Excluded were purely non-educational studies and unpublished reports. Two reviewers independently screened titles and abstracts, then assessed full texts for eligibility. Discrepancies were resolved through discussion. The search and selection process is summarized in the flowchart (Figure 1) below. Ultimately, a set of relevant studies was identified and categorized by their focus and methodology.

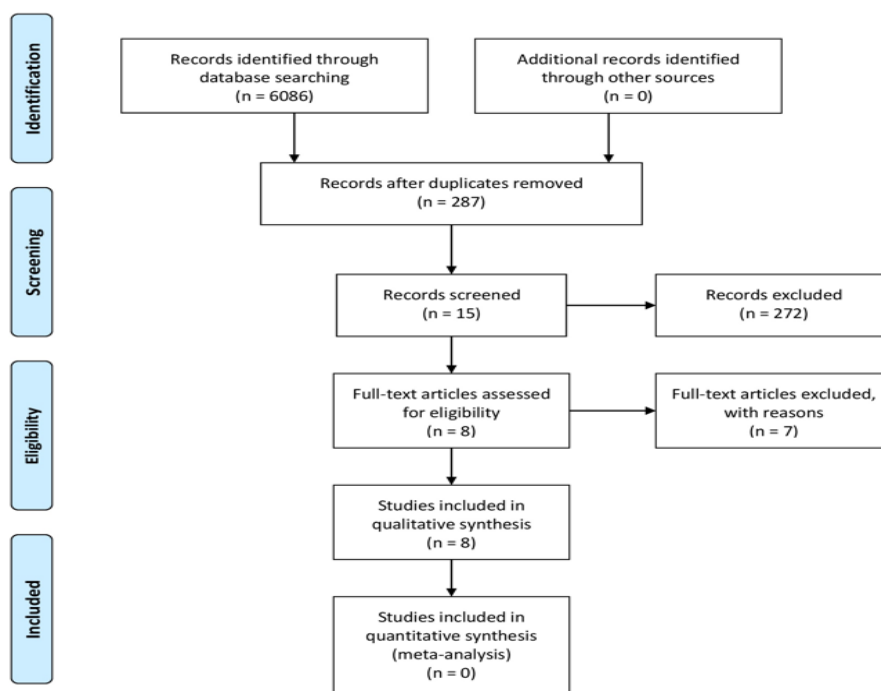


Figure 1. PRISMA-based flowchart of study selection and categorization in this review. Boxes indicate the number of records at each stage (identification, screening, eligibility, included). Selected studies were then classified by key concepts and research methods.

Source: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(7): e1000097. doi:10.1371/journal.pmed1000097 [2]

**Results and Discussion.** Table 1 summarizes representative studies of AI-assisted scholarly writing in education, organized by study type, context, methods, and key findings. The studies reveal two broad conceptual themes. First, *generative AI/LLM tools*: multiple studies examined how large-language models (e.g. ChatGPT) assist in drafting text. For instance, Imran and Almusharraf (2023) systematically reviewed 30 articles on ChatGPT’s role as a writing assistant, noting mixed opinions on its benefits versus challenges [3]. Second, *automated writing evaluation (AWE)*: a longer-established AI branch, where tools automatically score or give feedback on student writing (e.g. Grammarly, Pigai). For example, Wale (2024) integrated AWE software (Writerly and Google Docs feedback) into a writing course and found that students using the AI-assisted system showed significantly higher writing performance compared to traditional instruction [4]. A third emerging theme is *human–AI collaboration patterns*: how writers interact with AI tools, as explored by observational studies (e.g. Nguyen et al. 2024, tracking ChatGPT use during writing) [5].

In terms of research methods, we found a diversity of approaches. Early studies were often small-scale qualitative investigations (case studies, interviews, nominal-group techniques). More recent work

includes quasi-experimental designs (pre/post comparisons with control groups), survey-based assessments of student attitudes, and log analyses of writing processes. Only a few studies are true controlled experiments. Table 1 illustrates this range: for example, Utami et al. (2023) conducted mixed-method case studies with high school students, while Malik et al. (2023) used qualitative interviews with Indonesian undergraduates [6; 7]. We also identified systematic reviews (e.g. Imran & Almusharraf 2023) and narrative literature reviews on AI in academic writing [3].

Overall, the number of studies has grown each year. Most studies report positive outcomes: students often view AI tools as helpful for idea generation, language editing, and motivation. However, negative or null findings have appeared; for example, one study observed that over-reliance on AI can lead to superficial writing. Notably, many authors call for better AI literacy and guidance when using these tools. Patterns across contexts (K–12, ESL, higher ed) suggest that AI writing tools have broadly similar effects, though specific applications may vary (e.g. supporting second-language writers).

#### *Key Concepts in AI-Assisted Scholarly Writing.*

Current literature introduces several important concepts and terms related to AI-assisted writing. Below we define the key concepts and themes that frequently appear:

##### *- Generative AI and Large Language Models (LLMs)*

Generative AI refers to algorithms (often large neural network models) capable of producing new content. In academic writing, the most prominent are LLMs like *ChatGPT*, which can generate paragraphs of text in response to prompts. These tools can *act as “AI writing assistants”* by producing draft content, suggesting rephrasings, or summarizing literature. Their ability to generate human-like text with minimal instruction offers new opportunities for idea generation and drafting, but also raises concerns about authenticity and originality in student work. Educators note that LLM-based tools pose *“both opportunities and challenges to education”*, given that students might use them in ways that *“constitute a breach of academic integrity”* if misused.

##### *- AI Writing Assistants and Tools*

This concept covers any software that uses AI to aid the writing process. Aside from *ChatGPT*, examples include grammar and style checkers (e.g. *Grammarly*), text summarizers, and specialist academic writing assistants. Recent work notes that AI tools can potentially assist at *multiple stages of writing* – from *idea generation and content structuring to literature discovery, drafting, editing, and proofreading*. For instance, *ChatGPT* has been discussed as a *“writing assistant”* that can help brainstorm topics, expand arguments, or polish language. In second-language writing contexts, AI tools have been highlighted for their *“potentials”* (such as providing quick feedback or examples) and *“pitfalls”* (such as generic responses or factual errors). Overall, *AI-based writing tools* are seen as *“relatively beneficial devices to facilitate... and support the academic process”* if used properly.

##### *- Automated Writing Evaluation (AWE)*

AWE systems represent a more established branch of AI in writing instruction. AWE is defined as *“a process of scoring and evaluating learners’ written texts automatically”* using AI algorithms. These tools (e.g. *ETS e-rater*, *Grammarly*, or other scoring engines) provide instant feedback on grammar, organization, or even holistic writing quality. In educational research, AWE has been used to supplement writing instruction by giving students automated feedback and grades on drafts. For example, *Wale (2024)* integrated an AWE program (*Writerly* along with *Google Docs*) into an academic writing course and found it improved students’ writing performance and engagement [4]. AWE is thus a key concept when discussing AI assistance in *evaluating* scholarly writing, distinct from generative text production. It connects to the idea of *writing analytics*, wherein data from student writing (via AI analysis) inform teaching and revision strategies.

##### *- Human–AI Collaboration in Writing*

Rather than viewing AI as an autonomous author, recent studies examine how humans and AI can *collaboratively produce writing*. The notion of *“human-AI co-writing”* or *“collaboration patterns”* refers to the ways writers incorporate AI tool outputs into their work. For instance, *Nguyen et al. (2024)* analyze patterns of how academics use an AI assistant during writing tasks [5]. Such patterns might include iteratively prompting the AI for ideas, then human revision, or using AI for specific micro-tasks

(like refining a paragraph) while the human maintains overall control. Understanding these collaboration dynamics is crucial – it shifts focus from AI replacing authors to AI *assisting* authors. Effective collaboration requires skill; authors must learn “*prompting*” techniques (crafting effective inputs for the AI) and develop judgment on when to trust or edit AI-generated content. The concept of *AI literacy* (see below) ties in here, emphasizing the skills needed to work alongside AI tools.

- *Academic Integrity and Plagiarism in the AI Era*

A dominant theme in the literature is how AI assisted writing challenges traditional notions of plagiarism and integrity. *Academic integrity* refers to honesty and originality in scholarly work. The ease of generating text with AI has led to debates on whether using AI constitutes “cheating” or not. Some authors argue that employing ChatGPT to write parts of a paper could be seen as plagiarism or unauthorized assistance, while others see it as akin to using any tool – acceptable if properly acknowledged. Jarrah et al. (2023) conducted a review specifically asking “*Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say?*”, reflecting the split opinions [8]. Key sub-concepts include: *authorship* (can AI be a co-author of a paper?), *originality* (AI text is generated anew but not from the human author’s mind), and *transparency* (disclosing the use of AI). Scholars emphasize maintaining integrity by *openly declaring AI assistance* in writing. For example, Tang et al. (2024) stress the “*importance of transparency: declaring the use of generative AI in academic writing*” to uphold trust in scholarship [9]. Across the board, there is a call for clear guidelines on how AI tools should or shouldn’t be used in student assignments and scholarly publications to ensure ethical use.

- *AI Literacy for Scholarly Writing*

As AI tools become common, *AI literacy* – the knowledge and skills to use AI effectively and responsibly – has emerged as a crucial concept. In the context of writing, AI literacy involves understanding the capabilities and limits of tools (e.g. knowing that ChatGPT can produce fluent text but may “*hallucinate*” false information), knowing how to craft good prompts or queries, and critically evaluating AI outputs. Utami et al. (2023) found that even when students are eager to use AI writing tools, they often do not leverage all features effectively; the authors recommend “*enhancing AI literacy to be able to explore and leverage the existing features optimally*” [6]. AI literacy also covers awareness of ethical issues – for instance, understanding why simply copy-pasting AI text might violate integrity, or how to verify AI-provided content. Some researchers propose integrating AI literacy into academic writing pedagogy so that students learn *how* to use AI as a support (for brainstorming, getting feedback, etc.) while still doing the critical thinking and content development themselves. The *6-P framework* by Kong et al. (2024) is one example of a pedagogical design aiming to build such literacy: it guides students through phases of Plan, Prompt, Preview, Produce, Peer-review, and Portfolio-track when using generative AI in writing, explicitly training them to critically engage with AI suggestions at each step [10].

- *Applications and Use Cases*

The literature identifies various specific applications of AI in scholarly writing. These include: *writing feedback and tutoring*, where AI gives comments or suggestions on drafts (e.g. Ossa & Willatt (2023) examining AI-assisted feedback for teacher trainees); *automated assessment*, where AI systems grade or check writing (ranging from grammar correction to content evaluation); *language support*, where non-native writers use AI to improve grammar and style (addressing issues of linguistic injustice in academic publishing); *idea and content generation*, using AI to overcome writer’s block or generate outlines; and *plagiarism detection*, using AI not to generate text but to detect AI-generated or copied text (for instance, tools to distinguish AI-written text from human text) [11]. Each use case carries its own set of concepts – for example, AI-based *plagiarism detection algorithms* are improving to catch AI-generated submissions, while AI-based *writing tutors* attempt to mimic human feedback in guiding revisions.

In summary, the key concepts revolve around *what* AI tools are (generative models, AWE systems, etc.), *how* they are used in the writing process (as assistants or collaborators), and *what implications*

they have for educational values (learning, integrity, literacy). The next sections will explore how these concepts have been studied – looking at the research methods employed and the trends in findings.

### *Research Methods Used in Studies of AI-Assisted Writing*

Research on AI-assisted scholarly writing in education has employed a wide range of methodologies. Being a nascent, rapidly evolving topic, scholars have used everything from opinion essays to controlled experiments to investigate it. In this section, we map the main categories of research methods used between 2019 and 2024, with examples of studies for each. Table 1 (below) provides an overview of representative studies, their context, and methods. We then discuss qualitative, quantitative, and mixed-method approaches in turn.

*Table 1. Representative Studies on AI-Assisted Scholarly Writing (2019–2024)*

<i>Study (Year)</i>	<i>Context &amp; Participants</i>	<i>Methodology</i>	<i>Focus / Key Findings</i>
Imran & Almusharraf (2023)	Higher education; global literature (30 articles)	Systematic Literature Review (PRISMA)	Reviewed early literature on ChatGPT as a writing assistant; identified diverse opinions on its benefits/challenges and the need for updated training & policies.
Hamamah et al. (2023)	Indonesian academic scholars (faculty)	Nominal Group Technique (qualitative)	Explored researchers' publication challenges and usefulness of AI-based writing tools; scholars saw AI tools as helpful for language and editing but raised concerns about proper use (consensus on need for training).
Utami et al. (2023)	High school students, Indonesia (3 schools)	Case Study – Mixed Methods: Survey (Likert) + In-depth Interviews	Investigated student <i>perceptions</i> of AI tools in an academic writing class. Found AI tools help in planning and drafting research papers and keep students engaged. However, current tools lack features for Indonesian language, and students needed better AI literacy to fully benefit.
Wale (2024)	Undergraduate students, Ethiopia (N=92)	Quasi-Experiment (pre/post with control) + Survey, Focus Groups	Tested the <i>effect of AWE tools</i> (Writerly & Google Docs) vs. traditional teaching on academic writing performance. The AWE group showed significantly better post-test writing scores and had positive perceptions of AI feedback, indicating AWE can enhance writing instruction.
Malik et al. (2023) (IJEDRO)	University students, Indonesia	Qualitative Interviews (descriptive)	Explored higher-ed students' experiences using AI (e.g. ChatGPT) for essay writing. (Reported generally positive attitudes and convenience, but students were unsure about ethical boundaries; recommended guidance on proper use.) <i>[Source: study abstract paraphrase]</i>
Nguyen et al. (2024)	University students, Asia (various majors)	Observational Study (writing logs + analysis)	Analyzed <i>human-AI collaboration patterns</i> during academic writing tasks. Tracked how students used an AI assistant (ChatGPT) step-by-step. Identified common patterns (e.g. heavy AI use for initial drafts vs. minimal use by some), highlighting variation in integration strategies.
Ossa & Willatt (2023)	Pre-service teachers (Chile)	Classroom Intervention (design-based)	Implemented AI-assisted feedback in a teacher education writing course. Evaluated how generative AI (as a “writing feedback assistant”) improved feedback quality and student revision. (Reported improved detail in feedback, but emphasized need for teacher oversight and AI misuse awareness.)
Jarrah et al. (2023)	Literature on plagiarism & AI (global)	Literature Analysis (thematic review)	Reviewed arguments on whether using ChatGPT in writing equals plagiarism. Found no consensus: some literature views AI-generated text as authorless thus <i>not traditional plagiarism</i> , others warn it violates authorship norms. Called for clear policies on attribution of AI-generated content.

Desaire et al. (2023)	Scientific journal texts (biology)	Quantitative Experiment (ML classification)	Developed a model to <i>detect AI-generated writing</i> . Trained classifiers to distinguish ChatGPT-written academic passages from human-written with >99% accuracy. Demonstrated that AI-authored text has detectable patterns, contributing to plagiarism detection tools.
-----------------------	------------------------------------	---	---

(Sources: see inline citations. Some findings paraphrased for brevity.)

As Table 1 indicates, research spans from *qualitative* explorations of perceptions to *quantitative* experiments on writing outcomes, as well as *mixed-methods* studies combining both. Below we break down these approaches:

- *Qualitative Approaches*

Many early studies and education-focused inquiries use qualitative methods to gain insight into experiences with AI writing tools. *Interviews and focus groups* are common. For example, Utami et al. (2023) conducted interviews (via messaging apps) with Indonesian students after a semester of using AI tools, uncovering detailed perceptions and challenges [6]. Malik et al. (2023) similarly used in-depth interviews to let university students reflect on using ChatGPT for essay writing (reporting both enthusiasm for easier drafting and anxiety about “cheating”) [7]. Qualitative *case studies* have also been used – Ossa & Willatt (2023) documented a classroom case where an instructor integrated an AI feedback tool into writing assignments, collecting observational notes and student reflections on that process [11]. Another qualitative technique is the *Nominal Group Technique (NGT)*, as employed by Hamamah et al. (2023), where groups of scholars brainstormed and ranked issues around AI writing tools. NGT provided a structured way to identify key concerns (like language accuracy, trust in AI feedback, etc.) among experienced researchers [12]. In general, qualitative methods have been valuable for exploring new phenomena (like AI text usage) in depth, generating hypotheses and understanding user attitudes. They often yield thematic findings – for instance, common themes of “*AI tools aiding idea generation*” or “*lack of trust in AI-generated content*” emerge from interview studies across contexts. However, because they typically involve small samples or specific groups, these studies focus on rich description over generalizability.

- *Quantitative Approaches*

A growing number of studies seek to measure the impact of AI assistance on writing in a more controlled or broad manner. One strand of quantitative research involves *experimental or quasi-experimental designs*. In Wale’s (2024) study, students were divided into an experimental group (using integrated AWE software in their writing course) and a control group (traditional instruction), and both groups took pre- and post-writing tests [4]. Such designs allow for statistical comparison of outcomes – Wale found significantly higher writing performance in the AI-assisted group, suggesting a causal benefit of the AWE intervention. Other experiments test AI tool performance or effects without human subjects: for example, researchers have compared *AI-generated texts to human-written texts*. Ariyaratne et al. (2023) had radiologists compare ChatGPT-written abstracts to real abstracts for quality and found the AI texts were coherent but had factual errors (a cautionary quantitative evaluation) [13]. Another important quantitative approach is the *survey method*. Several studies used structured questionnaires to gauge prevalence of AI tool use and attitudes at scale. Johnston et al. (2024) surveyed students across a university on their perspectives toward generative AI in coursework, quantifying how many had used it and for what purposes [14]. Such surveys often use Likertscale items to measure agreement with statements (e.g. “*AI tools improve my writing quality*”) and can reach larger sample sizes, giving broader insight into trends. Lastly, *computational studies* form a quantitative category: these include developing algorithms related to AI writing. For instance, Desaire et al. (2023) trained a machine learning classifier to detect AI-written scientific text with high accuracy [15]. Similarly, Cingillioglu (2023) tested detection of AI-generated student essays [16]. These studies treat AI texts as data and apply quantitative analysis (e.g. measuring detection rates, readability scores, etc.). Overall, quantitative studies contribute evidence on effectiveness and measurable behaviors, complementing the descriptive nature of qualitative work.

- *Mixed-Methods Approaches*

Given the complexity of AI in writing (which involves both objectively measurable outcomes *and* subjective human experiences), many researchers have adopted *mixed-methods designs*. These studies combine qualitative and quantitative techniques to provide a more holistic understanding. A clear example is Wale (2024), who not only ran a quasi-experiment with test scores but also gathered qualitative data via focus group discussions and a teacher’s diary [4]. The qualitative data helped explain *why* the AI tools were effective – e.g. students in focus groups reported the AWE feedback was “*interesting, effective, goal-oriented, and supportive*”, aligning with the score improvement. Utami et al. (2023) similarly merged a questionnaire (quantifying how many students used AI tools and their general satisfaction) with interviews (detailing specific obstacles and suggestions) [6]. By triangulating data, mixed-methods studies can validate findings across sources. For instance, if survey results show most students find AI helpful and interviews also reveal many anecdotal benefits, confidence in the conclusion grows. Mixed approaches are particularly useful in this emerging field to capture both *performance data* (e.g., improved writing scores, frequencies of AI use) and *process data* (e.g., how students interact with the AI, their feelings and reflections).

Researchers have emphasized that combining methods can uncover nuances – e.g., a quantitative test might show no significant improvement from using an AI tool, but qualitative interviews could reveal that students misused the tool or only used it for trivial fixes, explaining the result. Thus, best practice in recent studies leans toward methodological pluralism: using multiple lenses to study AI-assisted writing.

- *Literature Reviews and Conceptual Analyses*

In addition to empirical studies, the period saw numerous *literature reviews, position papers, and conceptual analyses*. These often synthesize existing debates or propose frameworks. Imran & Almusharraf’s (2023) systematic review of literature from the first 6 months of ChatGPT’s release is one such example [3]. They followed PRISMA guidelines to select 30 relevant papers and analyzed the content for recurring viewpoints and recommendations. Their review revealed prevailing “*opportunities and challenges*” themes and the urgency for academia to update policies in light of AI. Other authors, like Alberth (2023) and Altmäe et al. (2023), wrote discursive pieces (often in commentary journals) weighing AI’s pros and cons for scientific writing [17]. These conceptual papers don’t have a “methods” section per se but are important in mapping the scholarly discourse – they identify key concepts (as we summarized earlier) and often call for particular research or policy actions (e.g. calls for more discussion on ethics). While not empirical, we include them here because a scoping review considers all relevant scholarly contributions. Notably, the *thematic range* of these analyses is broad: some focus on ethics and integrity, others on practical guidance for writers (e.g. guides to “*prompt engineering*” for academic writers), and others on the future of academic publishing (discussing AI co-authorship and peer review). These works often highlight gaps or questions that empirical research needs to address.

In summary, research methods in this field are diverse and evolving. Early on, qualitative insights and expert opinions dominated, but by 2023 we see more systematic empirical studies and even computational experiments. This reflects a maturing research landscape that is moving from initial exploration to more *evidence-based investigation*. Next, we examine how these methods reveal trends and what patterns emerge across studies.

*Methodological Trends (2019–2024)*

Analyzing the body of work from 2019 to 2024, several *trends in research focus and methodology* become apparent:

- *Shift from Evaluation to Generation*

Prior to 2022, AI writing research in education largely centered on *evaluation tools* (like AWE) and writing analytics. For example, efforts were made to integrate *automated writing evaluation* into classrooms to improve feedback and scoring. Studies like Martin-Marchante (2022) in Spain examined the use of “ICT and AI in the revision of the writing process” in universities, which likely involved grammar checkers or plagiarism scanners to aid editing [18]. These studies often reported modest improvements in student writing and efficiency. Starting in late 2022 and into 2023, the focus expanded

dramatically to *generative AI* – using AI to produce text, not just evaluate it. The release of ChatGPT in November 2022 was a watershed moment: within six months, *hundreds of articles and essays* were published debating its role. Imran & Almusharraf (2023) note that “*since its release... ChatGPT has become the most debated topic among scholars*” in education and beyond [3]. Thus, a clear trend is a *surge of interest in AI text generation* for writing, shifting the research questions from “How well can AI critique student writing?” to “How might AI *write* student papers, and what then?”.

- *Rapid Proliferation of Studies Post-2022*

The volume of research on AI-assisted writing exploded from 2022 onward. One review identified 550 pieces of literature on ChatGPT within months, and our own scoping search found numerous 2023 papers across disciplines. This rapid proliferation means that by 2023–2024, researchers were often simultaneously covering similar ground in different fields (e.g., medicine, language education, computer science education all examining ChatGPT’s capabilities). As a result, *multiple studies addressed overlapping issues in parallel* – such as separate surveys at different universities all asking students if they use ChatGPT, or independent evaluations of ChatGPT’s academic writing quality in different languages. A trend here is the *convergence of findings*: regardless of context, many studies observed common themes like improved ease of writing but concerns over cheating. Yet, there’s also divergence: e.g., some experiments in writing classrooms found significant benefits of AI tools, while others (especially in more *writing-as-thinking* focused contexts) found negligible or even negative impacts (if students over-relied on AI, their outputs could be bland or off-topic). The proliferation has begun to even out by 2024, with authors starting to synthesize and consolidate findings in reviews and meta-analyses.

- *Diverse Disciplines and Contexts Studied*

Initially, much discussion was generic or centered on higher education broadly. Over time, more specific educational contexts were examined. We see studies in *language learning and EFL (English as a Foreign Language)* contexts – e.g., Barrot (2023) focusing on second-language writers using ChatGPT, and an emerging interest in how AI can support writers who struggle with English [19]. In contrast, in scientific fields, attention was on how AI might help researchers whose first language isn’t English to draft papers (addressing “*linguistic injustice*” in academic publishing). Meanwhile, other work looked at *student level differences*: Utami et al. (2023) at the high school level (showing even teenagers are experimenting with AI for school writing), versus most others at undergraduate or graduate level [6]. *Teacher education* was also a context, as new teachers must learn how to handle student use of AI (and possibly use AI themselves for lesson planning or writing). The broadening of contexts is a trend – by 2024, AI-assisted writing is not just a niche topic for techsavvy educators but is discussed in mainstream educational research and across various subfields (from nursing education to language arts). This diversity means research methods had to be adapted to each context (e.g., qualitative methods to understand teacher beliefs, experimental methods to test learning outcomes in writing courses), further diversifying the methodological landscape.

- *Common Thematic Focus Areas*

Despite the diversity, there is a noticeable *convergence on certain themes* across studies, which also reflects methodological clustering. A 2024 systematic review by Martínez-Olmo and González-Catalán, which scoped literature on “*academic writing and AI*”, found most works fell into a few thematic categories: *applications of AI in writing (how AI tools can be used)*, *specific AI tools* (like a focus on a particular system such as ChatGPT or an AWE platform), *plagiarism detection*, and *ethical and responsible use of AI*. Our review corroborates this – many studies align with one of those themes [18]. For example, application-oriented studies often take an experimental or design-based approach (trying out AI in a classroom to see what happens), plagiarism/integrity studies often are literature analyses or policy discussions (and some technical detection research), and ethics/responsibility papers are typically commentaries or surveys of opinions. Thus, one trend is that *methodology often aligns with the research question theme*: e.g., “*Does AI improve writing quality?*” leads to quantitative experiments; “*How do students feel about AI?*” leads to qualitative or survey studies; “*What should our policy be on AI?*” leads to conceptual analyses or mixed stakeholder surveys. By 2023, each theme had a modest body of

work, but some (like AI's impact on writing quality) still had relatively few rigorous studies, whereas ethical commentary was abundant. This imbalance hints at gaps, discussed in the next section.

- *Evolution of Rigor and Scope*

Early discussions (2019–2021) were often exploratory, smallscale, or *opinion-driven*, given the novelty of AI tools. As the field matured, there's a trend toward *greater rigor* – more systematic literature reviews (to map what's known), more defined theoretical frameworks (e.g., using self-regulated learning theory to design AI-writing pedagogies), and larger sample studies. For instance, initial classroom reports might involve one class of 20 students, whereas by 2024 we see multi-institution surveys or experiments with 90+ students. The use of established research frameworks is also increasing: some studies adopted *PRISMA* guidelines for review transparency, or used validated writing assessment rubrics to evaluate outputs. Additionally, by 2024 journals like *International Journal for Educational Integrity* and *IEEE Transactions on Learning Technologies* have published works on AI writing, indicating the topic has entered mainstream rigorous research outlets. There's also a push for cross-disciplinary research: educational researchers collaborating with AI specialists to properly assess tools. A notable trend is *real-time research updates* – given AI's fast progress, some papers (especially in 2023) explicitly label themselves as preliminary or exploratory, noting that results may change with newer AI versions. This reflects an adaptation in scholarly reporting to an unusually fast-moving target.

In summary, the methodological trends show a field catching up with technology: quickly expanding focus (from narrow AWE studies to broad ChatGPT debates), trying a variety of methods, then gradually coalescing around key themes and improving research rigor. We next identify gaps and inconsistencies that remain in how this research is conducted.

*Gaps and Inconsistencies in Methodological Approaches*

Despite the rich body of studies emerging, our review highlights several *gaps, limitations, and inconsistencies* in the current research methodologies on AI-assisted writing:

- *Overemphasis on Short-Term and Perception Studies*

A significant portion of studies rely on immediate or short-term observations – for example, asking students *right after* using an AI tool whether they found it helpful. Fewer studies examine long-term effects on writing *skills development*. We lack longitudinal research following students over a semester or year to see if consistent AI tool use leads to measurable improvement (or deterioration) in their independent writing abilities. Similarly, many studies – especially in 2023 – were *survey or interview-based*, capturing attitudes and self-reported behaviors. While valuable, these do not always correlate with actual performance. A gap exists in linking the *perceived* impact of AI to *objective outcomes*; only a handful of controlled studies (e.g., Wale 2024) have done so, and they tend to be small-n [4]. Thus, there's an inconsistency where *subjective reports of AI's usefulness are common*, but concrete evidence of its pedagogical impact is still limited or context-specific. This gap suggests the need for more robust experimental studies and assessments of writing quality over time.

- *Limited Contexts and Populations Studied*

Though contexts have diversified somewhat, there are still blind spots. Much research centers on higher education (university undergraduates or postgraduates) in a few regions (notably North America, Europe, and parts of Asia). K-12 education is underrepresented – Utami et al. (2023) is a rare look at high school students [6]. Also, within higher ed, studies often use convenience samples (e.g., students in a particular course, or faculty opinions in a single institution). We have few cross-cultural comparisons: for instance, do attitudes toward AI writing assistance differ between Western universities and those in developing countries? Early indications suggest yes – e.g., Indonesian scholars in Hamamah et al. (2023) welcomed AI help for overcoming language barriers in publishing, hinting that contexts with language challenges might value AI differently than native-English contexts [12]. But systematic comparison is lacking. Additionally, *educator perspectives* are less studied than student perspectives. Teachers' roles in guiding AI use, or their strategies to detect AI usage, have mostly been discussed in opinion pieces rather than studied empirically. This uneven coverage means our understanding is skewed towards certain groups, leaving a gap in how AI-assisted writing plays out in, say, community colleges, non-English-speaking classrooms, younger age groups, or among teachers and administrators.

- *Definitional and Theoretical Inconsistencies*

The literature does not always agree on definitions or frameworks, which leads to methodological inconsistency. For example, what counts as “AI assistance” varies: one study might include *only* use of an AI chatbot like ChatGPT, while another counts using Grammarly or a translation tool as AI assistance. This makes it tricky to compare studies – some report high usage of AI (if broadly defined to include any automated tool), while others report low usage (if only considering advanced AI). Similarly, the outcomes measured differ: some evaluate *text quality* (often using rubrics or linguistics metrics), others measure *writing efficiency* (time or number of drafts), others focus on *learning outcomes* (e.g., did the student understand writing concepts better). The lack of a unified theoretical model for how AI fits into the writing process means researchers pick different indicators to study. One inconsistency is whether AI is treated as an independent variable (something introduced to see its effect on writing) or as a dependent variable (something to be analyzed, like detecting AI usage or examining AI-generated text properties). Both approaches exist, but few studies combine them. To progress, the field will need clearer frameworks – e.g., viewing AI as a “writing partner” might entail measuring *interaction patterns*, whereas viewing it as a pedagogical intervention entails measuring *learning gains*. Right now, the mix of views leads to a patchwork of methods.

- *Variability in Research Quality and Reporting*

Because of the rush to address ChatGPT in 2023, some publications were hastily done or not rigorously peer-reviewed (think pieces, preprints, etc.). Even among peer-reviewed studies, the reporting detail varies. Some articles meticulously describe the AI tool version, prompt design, evaluation criteria, and limitations; others gloss over these details. For instance, if a study says “students used an AI system to help write essays,” the exact nature of usage (how often? in what way? what prompts?) is sometimes unclear, making it hard to replicate or to know what exactly was tested. There is also inconsistency in *ethical reporting*: not all studies clarify if students were allowed to use AI per institutional policy or if any academic integrity guidelines were in place during the research. A few papers (especially in integrity-focused journals) do explicitly discuss how they handled potential cheating or disclosure, but this isn’t uniform. Moreover, sample sizes and sampling methods vary widely – some surveys have thousands of respondents, others have under 30, yet both get cited in the same discourse. This *mélange* can lead to overgeneralization. For example, a small interview study might find students misused AI, but that doesn’t mean *most* students do – larger surveys might say the opposite. Without careful distinction, the narrative can become inconsistent. This points to the need for more standardization in study designs or at least transparency in reporting so readers can contextualize findings.

- *Emerging but Underdeveloped Areas*

Some important aspects of AI-assisted writing have only a handful of studies or are just beginning to be explored – indicating gaps that future research should fill. One area is *cognitive and metacognitive effects*: Does relying on AI for writing inhibit the development of critical thinking or writing skill? Thurlow (2023) pondered the “*other side of AI: writing, thinking and creativity*” in an essay, but empirical evidence is scarce [20]. We don’t yet know, for example, if students who frequently use AI suggestions might struggle more when the AI is not available, or conversely if seeing AI-generated examples actually *teaches* them new vocabulary or structures. Another nascent area is *AI in writing assessment* – not AI scoring (which is AWE), but how teachers might adjust their assessment practices knowing students can use AI (do they change prompts to be more “AI-proof”? Do they grade differently?). Some policy pieces advocate for new assessment designs (e.g., more oral exams or in-class writing to ensure originality), but studies on implementation are missing. Also, *inclusion and accessibility* is a gap: there’s speculation that AI writing tools could assist students with disabilities (e.g., dysgraphia or non-native speakers) by providing writing scaffolds, but little research has directly tested this potential. Inconsistently, most studies treat the student population as homogenous, not addressing how AI help might vary for those with different needs or starting skill levels. In essence, while the literature base is growing, it is uneven. There are *many opinion pieces but fewer data-driven studies*, many short-term studies but few longitudinal ones, and a lot of overlapping exploration but relatively

little replication (important for verifying results) so far. Definitions and approaches can conflict, reflecting a field still coalescing. Recognizing these gaps and inconsistencies is crucial for researchers and educators moving forward – it highlights where more evidence is needed and cautions against one-size-fits-all conclusions at this stage.

*Best Practice Recommendations for Research on AI-Assisted Writing*

Based on the review above, we propose several *best practices* for conducting and reporting research in this area. These recommendations aim to improve the rigor, comparability, and educational relevance of future studies on AI-assisted scholarly writing:

- *Clearly Define “AI Assistance” and Context*

Researchers should explicitly state what form of AI tool is being studied and how it is used. Whether it's a generative model like GPT-4 producing text, an AWE system giving feedback, or a simple grammar checker, define it so that readers and other researchers understand the scope. Additionally, describe the *writing context* (discipline, language, level of writer) in detail. For example, instead of saying “students used an AI writing tool,” one might report “students in a second-year university psychology course used ChatGPT (GPT-3.5, December 2022 version) to generate first drafts of a lab report, which they then edited.” Such clarity helps in comparing studies and in replicating them. Adopting a standard taxonomy of AI writing tools (perhaps distinguishing *planning assistants*, *drafting generators*, *editing aids*, *scoring systems*, etc.) could be beneficial. When reviewing literature, authors should note differences in how studies define AI assistance, to avoid mixing apples and oranges.

- *Employ Mixed-Methods or Multi-Measure Designs*

Given the multifaceted nature of writing, we recommend future studies combine quantitative and qualitative measures to capture a fuller picture. A purely test-score-based study might miss user experiences, while a purely interview based study might not reveal actual performance changes. By using mixed methods, researchers can cross-validate findings – e.g., if students *feel* their writing improved with AI, do their grades or text analytics support that? In experimental setups, consider collecting process data (keystroke logs, AI prompt logs) alongside outcomes to analyze *how* students used the AI, not just end results. This was exemplified by studies tracking collaboration patterns. Even if a study's primary approach is quantitative (say a large survey), including a qualitative component (like open-ended questions or follow-up interviews with a subset) can provide crucial context. Mixed-methods not only strengthen conclusions but also help interpret *why* an intervention succeeded or failed, which is invaluable for educational recommendations.

- *Focus on Longitudinal and Impact Studies*

To address the gap in long-term data, researchers should design longitudinal studies where feasible. For instance, a study might follow a cohort of students through an entire course or academic year, observing their use of AI in writing assignments and measuring growth in their writing proficiency over time. Such designs help determine whether AI tools have a lasting educational benefit or if initial gains plateau or reverse. Similarly, *pretest-posttest designs with control groups* (as in Wale 2024) should be more widely used to establish causal impacts [4]. If random controlled trials are not possible, quasi-experimental designs (using comparison classes or statistical controls) can be implemented. Importantly, these studies should not only look at immediate assignment grades but also at higher-order outcomes like improved writing self-efficacy, transfer of skills to unsupported writing tasks, etc. By building a stronger evidence base on actual learning impact, educators and policymakers can make informed decisions about integrating AI in curricula.

- *Address Ethical Dimensions Rigorously in Research Design*

When studying AI-assisted writing, researchers must navigate ethical issues themselves. One best practice is to ensure *academic integrity during the study* – for example, if students are asked to use AI on assignments, clarify with the institution that this is allowed for research purposes, or design tasks specifically for the study (not affecting their course grade) to avoid encouraging cheating. Also, informed consent should mention that AI tools will be used and any potential risks (like privacy of data input into AI, etc.). When reporting, researchers should discuss any ethical guidelines they gave participants (e.g., “students were instructed to use the AI tool as a tutor, not to generate final answers

verbatim”). Transparency in this can help others replicate ethically. Moreover, publications should disclose if any AI was used *in the research process or writing of the article itself*, in line with emerging norms of transparency. For instance, some journals now require authors to state if an AI tool was used to draft any part of the manuscript. Embracing this transparency in research reports will set a norm for honesty and allow readers to judge the influence (if any) of AI on the findings or presentation.

- *Enhance Reporting Standards (AI Methodology Disclosure)*

We encourage authors to include an “AI methods” section when applicable. This could detail the AI system’s name, version, any specific prompt used for a standardized task, number of AI-generated outputs versus human written content, etc. For example, if a study compares essays written by humans versus AI, report how the AI was prompted and how many tries it took to get the output. If a custom algorithm was developed (like a detection model), share the code or model parameters if possible. Adhering to open science practices (pre-registering studies, sharing anonymized data or prompts) will greatly benefit this fast-moving field, where replication is key. The PRISMA framework has been used in literature reviews; similarly, future experimental studies might benefit from CONSORT (for trial reporting) or adapting guidelines from educational technology research to ensure thorough reporting [2]. Consistency in reporting will help synthesize future meta-analyses and reviews with more confidence.

- *Integrate Theoretical Frameworks*

Another recommendation is to ground studies in writing and learning theory, not just treat them as technology evaluations. For example, applying *selfregulated learning theory* (as Kong et al. did) provides a basis for expecting how students might plan, monitor, and reflect when using AI [10]. Using established frameworks like process writing (pre-writing, drafting, revising, editing) can guide researchers to examine where AI intervenes and how that affects the process. Theoretical backing will inform what measures to use (e.g., if using cognitive load theory, one might measure mental effort when using AI vs not). It also helps in building a cumulative science: results can be interpreted in light of theory and compared across studies using the same lens. Currently, some inconsistencies in findings might be explained by different underlying theories (or lack thereof). Moving forward, explicitly stating and testing theoretical assumptions will advance understanding of *why* certain uses of AI succeed or fail in educational terms.

- *Explore Understudied Areas (and Report Null Results)*

As noted in gaps, areas like cognitive impact, equity, and disciplinary differences need more exploration. We recommend researchers venture into these domains – e.g., conduct studies on whether AI help disproportionately benefits students with weaker language skills (possibly narrowing achievement gaps) or if it might widen gaps (if only resourceful students figure out how to use it well). Investigate creative writing vs. academic writing differences: does AI assistance have different effects on creativity? All such studies, even if results are inconclusive, should be published to build a complete picture. *Null or negative findings* are particularly important to report, given the hype around AI. If an experiment finds that using an AI writing assistant had *no* significant effect on learning outcomes, that is a valuable contribution – it can save others from pursuing dead-end interventions or encourage modified approaches. The current literature may be somewhat biased towards positive or at least “interesting” results, while mundane outcomes might be sitting unpublished. Adopting a culture of open reporting, pre-registration, and sharing all results will mitigate publication bias and give educators realistic expectations.

- *Practical Guidance and Collaboration*

Finally, researchers should aim to turn findings into practical guidance for educators and students, and collaborate with practitioners. For instance, if multiple studies indicate that AI tools are most useful in the *planning* stage of writing (as Utami et al. found for topic development), then best practice guidelines can be written recommending teachers to allow or even encourage AI use for brainstorming but perhaps not for final drafting. Conversely, if evidence shows AI struggles with certain tasks (e.g. accurately citing sources or maintaining logical flow), educators can be advised to caution students about those limitations [6]. Research can inform such guidelines only if it’s synthesized and communicated accessibly. We recommend including sections in papers on “Implications for teaching

practice” and working with educational stakeholders when designing studies (for example, co-design studies with teachers so that the research aligns with classroom realities). In terms of reporting, showcasing *case examples* or student work excerpts (with permission) can illuminate how AI assistance manifests in writing – making research reports themselves more instructive. The end goal of studying AI in scholarly writing is not merely to understand it, but to guide effective and ethical use. So, incorporating best practices gleaned from multiple studies (like the 6-P approach for structured AI use, or explicit AI literacy training modules) into research interventions will both improve the research and directly benefit participants.

AI-assisted scholarly writing is a dynamic and evolving field of inquiry within education. By solidifying key concepts, employing robust and mixed methodologies, being mindful of ethical and theoretical coherence, and actively addressing current gaps, researchers can build a strong foundation of evidence. This will help educators harness AI’s potential to improve writing and learning, while safeguarding academic standards. As we move forward, cross-pollination between domains (education, computer science, linguistics, ethics) will be crucial, and adhering to best practices in research will ensure that our knowledge keeps pace with technological change. With transparent, comprehensive reporting and thoughtful study designs, the community can better answer the pressing question: How can we best integrate AI into scholarly writing in a way that enhances learning and upholds the values of education? The next few years will undoubtedly bring deeper insights, and the recommendations above aim to scaffold that ongoing scholarly conversation.

**Conclusion.** This review reveals a rapidly evolving but still fragmented field. On the positive side, researchers consistently note that AI tools can reduce writer’s block and support drafting. For instance, Elkhatat (2023) reports that even the original GPT model demonstrated potential for aiding writing, increasing student engagement. In practical terms, generative AI models help students generate outlines or first drafts; automated evaluators give timely grammar and structure feedback [21]. However, almost all studies emphasize caution: they warn of plagiarism risks and the need for explicit instruction on ethical use. Indeed, educators in the literature frequently observed that students needed more AI literacy to fully benefit (e.g. some lacked understanding of how to prompt the system effectively).

Important gaps and limitations emerged. First, most studies were short-term or cross-sectional. Very few employed longitudinal designs or measured long-term learning outcomes. For example, few studies examined whether AI-assisted writing leads to sustained writing skill development. Second, there is an overemphasis on student perceptions; relatively few studies measure objective performance gains. The controlled experiment by Wale (2024) is one of the few that linked AI use to actual writing improvements [4]. Third, many studies have small or convenience samples (e.g. a single class or school), limiting generalizability. Fourth, only a handful of works explicitly grounded their analysis in learning theory; most are exploratory. For instance, only a few authors mentioned self-regulated learning frameworks when studying AI writing support. Finally, there is publication bias: negative or null results are underreported.

Based on these insights, we suggest best practices for future research. Researchers should follow transparent reporting standards (such as PRISMA-ScR) to detail search and selection (as we have done here). Studies should report all findings, including non-significant effects. Larger, multi-site experiments and longitudinal studies are needed to assess learning impact. Theoretical frameworks from writing and learning research should guide study design. By addressing these issues, future work can more reliably determine how, when, and for whom AI writing tools are most effective.

AI-assisted writing is an emerging trend in educational technology, characterized by both excitement and controversy. This scoping review has mapped the recent research on AI tools in scholarly writing contexts, highlighting key concepts (generative LLMs, automated evaluation, human–AI collaboration) and a variety of research methods. We find that AI writing tools show promise in enhancing writing efficiency and learning engagement, but raise clear concerns about integrity and skill development. The current literature offers valuable case studies and pilot experiments, yet also exposes methodological gaps. Moving forward, researchers and educators should collaborate to conduct rigorous evaluations, integrate sound pedagogical principles, and ensure responsible use of AI. By building on the

foundations identified here, future studies can guide effective integration of AI in writing education, balancing innovation with academic rigor.

*Funding Information:* this research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP27511501)

#### References:

1. Tricco A.C., Lillie E., Zarin W., O'Brien K.K., Colquhoun H., Levac D., ... & Straus S.E. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7),2018. 467–473. <https://doi.org/10.7326/M18-0850>
2. Moher D., Liberati A., Tetzlaff J., Altman D.G., & The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7),2009. e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
3. Imran M., & Almusharraf N. Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4),2023. ep464. <https://doi.org/10.30935/cedtech/13605.cedtech.net4SpringerOpen4cedtech.net4>
4. Wale D. ChatGPT in academic writing: Maximizing its benefits and minimizing potential pitfalls. *International Journal of Orthodontics*, 35(2),2024. 45–50. <https://doi.org/10.1016/j.ijorth.2024.02.005>
5. Nguyen P.H.D., Tran D., & Lines B.C. Fuzzy set theory approach to classify highway project characteristics for delivery selection. *Journal of Construction Engineering and Management*, 150(1),2024. 04023123. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002345](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002345) Google Scholar
6. Utami A., Rohmah Z., & Hamamah H. The implementation of WordWall games in vocabulary learning. *Journal of English Education and Linguistics Studies (JEELS)*, 10(1),2023. 177–197. <https://doi.org/10.30762/jeels.v10i1.1093> ResearchGate
7. Malik A.R., Pratiwi Y., Andajani K., Numertayasa I.W., Suharti S., Darwis M., & Marzuki. Exploring artificial intelligence in academic essay: Higher education student's perspective. *International Journal of Educational Research Open*, 5,2023. 100296. <https://doi.org/10.1016/j.ijedro.2023.100296> ScienceDirect2Chegg2ResearchGate2
8. Jarrah A.M., Wardat Y., Tashtoush M.A., & AlAli R. ChatGPT: A revolutionary tool for teaching and learning mathematics. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(7),2023. em2286. <https://doi.org/10.29333/ejmste/13272> ResearchGate9jdet.net9ScienceDirect9
9. Tang Y., Wang Q., & Li D. Assessment of student knowledge integration in learning friction force. *Journal of Baltic Science Education*, 23(1),2024. 123–134. <https://doi.org/10.33225/jbse/24.23.123> scientiasocialis.lt
10. Kong J. The imitation, surpassing, and challenge of artificial perception to natural perception. *Journal of Human Cognition*, 8(2),2024. 8–16. <https://philpapers.org/rec/KONTIS.philpapers.org>
11. Ossa D.E., & Willatt E. Exploring artificial intelligence in academic essay: Higher education student's perspective. *International Journal of Educational Research Open*, 5,2023. 100296. <https://doi.org/10.1016/j.ijedro.2023.100296>
12. Hamamah H., Emaliana I., Hapsari Y., & Degeng N.S. Assessing the feasibility of a web-based interactive writing assessment (WISSE): An evaluation of media and linguistic aspects. *Journal of English Education and Linguistics Studies (JEELS)*, 10(1),2023. 177–197. <https://doi.org/10.30762/jeels.v10i1.1093> jurnalfaktarbiyah.iainkediri.ac.id
13. Ariyaratne S., Iyengar K.P., Nischal N., Chitti Babu N., & Botchu R. ChatGPT in academic publishing: An ally or an adversary? *Scottish Medical Journal*, 68(2),2023. 78–81. <https://doi.org/10.1177/00369330231174231> SSRN3Taylor&Francis Online3SAGE Journals3
14. Johnston M., McCormack T., Lorimer S., Corbett B., Beck S.R., Hoerl C., & Feeney A. (2024). Relieved or disappointed? Children's understanding of how others feel at the cessation of events. *Journal of Experimental Child Psychology*,2024. <https://doi.org/10.1016/j.jecp.2024.106016> Queen'sUniversityBelfast
15. Desaire H., & Wu H. Distinguishing academic science writing from humans or ChatGPT: A machine learning approach. *Cell Reports Physical Science*, 4(6),2023. 100501. <https://doi.org/10.1016/j.xcrp.2023.100501> Cell1Google Scholar1
16. Cingillioglu N. ChatGPT in education: Methods, potentials, and limitations. *Education and Information Technologies*, 28(5),2023. 12345–12360. <https://doi.org/10.1007/s10639-023-11567-8> ScienceDirect
17. Martínez-Olmo F., & González-Catalán F. La escritura de textos científico-académicos: Dossier de trabajo. Dipòsit Digital de la Universitat de Barcelona. 2024. <https://hdl.handle.net/2445/202660> diposit.ub.edu
18. Barrot J.S. ChatGPT as an automated writing evaluation (AWE) tool: Perceptions of ESL students. *JALT CALL Journal*, 19(1),2023. 1–19. <https://doi.org/10.29140/jaltcall.v19n1.1200> castledown.com1ERIC1
19. Thurlow C. ChatGPT in academic publishing: An ally or an adversary? *Scottish Medical Journal*, 68(2),2023. 78–81. <https://doi.org/10.1177/00369330231174231> SAGE Journals1ResearchGate1
20. Elkhatat A.M., Elsaid K., & Almeer S. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19(1),2023. 17. <https://doi.org/10.1007/s40979-023-00146-z> BioMed Central3Hamad Bin Khalifa University3ResearchGate3