

“A Systematic Review of AI Tools Supporting Academic Research Writing in Saudi Arabia”

Researcher:

Lmya Hussain A Al-Zahrani

MBA, KAU, jeddah, Saudi Arabia.



<https://doi.org/10.36571/ajsp8520>

Abstract:

This systematic review assessed the peer-reviewed literature on artificial intelligence (AI) tools in supporting academic research writing within Saudi Arabia and evaluated their benefits, challenges, and gaps. Major databases (e.g., Scopus, Web of Science, IEEE Xplore, ERIC, Google Scholar) were searched utilizing Boolean combinations related to AI, academic writing, and Saudi higher education. Screening followed PRISMA 2020, and methodological quality was appraised using CASP checklists. Of 950 records, 19 studies met the inclusion criteria. Most included studies were cross-sectional or experimental and focused on undergraduate and postgraduate EFL/ESL learners. Tools such as ChatGPT, Grammarly, QuillBot, and Turnitin were most commonly used, primarily supporting and post-writing tasks (e.g., paraphrasing, grammar correction, fluency enhancement, plagiarism checks). By contrast, pre-writing discovery tools (e.g., Elicit, Scite, Consensus) and referencing/management utilities (e.g., citation managers) received relatively limited attention, indicating a surface-oriented adoption pattern. Reported benefits centered on improved linguistic accuracy, time efficiency, and heightened student confidence. Supervisors noted that delegating mechanical corrections to AI freed time for higher-order feedback on structure, argumentation, and methodology. Nevertheless, risks were recurrent: plagiarism and authorship ambiguity, reduced critical thinking stemming from overreliance on automated outputs, inconsistent disclosure policies, and infrastructural and digital-literacy disparities that constrain equitable adoption. General, the evidence indicates that AI tools function best as supplements rather than substitutes for academic writing. A balanced, policy-aligned model, emphasizing training, supervisory mediation, transparent disclosure, and equitable access is essential for sustainable integration aligned with Saudi Arabia's Vision 2030.

Keywords: Artificial intelligence; Academic writing; ChatGPT; Grammarly; Saudi Arabia; Systematic review.

1. Introduction

Artificial Intelligence (AI) has revolutionized academic writing research in numerous countries, originating from the periphery to a complete set of software for drafting, revising, and determining integrity. Through the academic treatises and writing, we make an important contribution to the field of scholarly communication through which scholars are able to input original information, announce results, and play an influential role in key discourse fields of world knowledge. The writing of the academic research text, whether an individual article or the thesis/dissertation, requires expertise not only in content knowledge but also a discrete competence in methodological procedures, and rhetorical strategies, as well as language use (Cheong et al., 2023; Teng & Wang, 2023). For many learners, especially in English as a Foreign Language (EFL) contexts such as Saudi Arabia, for example, this has led to persistent problems related to grammar, cohesion and coherence, argumentation and academic integrity (Aljuaid 2024; Abduljawad 2024).

In this scenario, artificial intelligence (AI) techniques have risen to prominence in recent years. The origins of AI in writing support (such as automated grammar checkers and plagiarism detection) have morphed into a wider ecology of natural language processing (NLP) based tools and machine learning. For example, the generalizable language models (LLMs) like GPT-3 (Brown et al., 2020) and GPT-4 (e.g., ChatGPT; Alharbi et al., 2024) allow for sophisticated draft writing, rewriting, and ideation. Applications like Grammarly and QuillBot target improving surface-level fluency and coherence, while tools like Elicit [14], Scite [1], and Consensus [44] focus on literature discovery and evidence synthesis. Both in combination cover the whole process of the research-writing pipeline from pre-writing to revision and integrity (Alanazi et al., 2025).

Evidence from international studies reveals that AI tools contribute positively to writing fluency, linguistic accuracy, and productivity (Perkins, 2023). They support decrease writer's block, accelerate drafting, and provide real-time feedback for revision (Alyami et al., 2025). For EFL researchers, the benefits are especially pronounced: Grammarly and QuillBot have been shown to improve grammar, reduce stylistic errors, and increase confidence in academic expression (El-Garawany, 2024; Fan, 2023). Yet, these gains are tempered by widespread concerns. Global systematic reviews indicate that while AI tools enhance surface-level quality, they do not necessarily foster originality, critical thinking, or methodological rigor (Hamoda et al., 2025; Deep & Chen, 2025). Risks include plagiarism, overreliance on machine-generated text, and the erosion of authentic scholarly voice (Alanazi et al., 2025). Ethical ambiguity, particularly around authorship and disclosure of AI use, remains unresolved in many contexts (Alharbi et al., 2024).

Saudi Arabia is undergoing a rapid transformation of its educational and research systems under the Vision 2030 framework, which identifies human capital development, digital innovation, and global research competitiveness as strategic priorities (Saudi Vision 2030; Ryan, 2023). Policy initiatives emphasize strengthening graduate education, expanding international

publishing, and embedding digital technologies across academic practice (Mohiuddin, 2023). In this environment, AI tools are increasingly viewed not only as linguistic aids but also as potential drivers of research productivity and international visibility.

Saudi universities have begun to integrate AI technologies into classrooms, research training, and institutional infrastructure. E.g., Ozfidan et al (2024) found that undergraduate students in Saudi Arabia frequently relied on ChatGPT, Grammarly, and Google Translate for writing assignments. Their factor analysis revealed two latent dimensions: “instructional support” and “instructional practices” of AI, which together explained more than 55% of variance in student perceptions. Abduljawad (2024) presented statistically significant progress in the writing performance of Saudi ESL students using AI revision and feedback tools. Alhusaiyan (2025) also mentioned opportunities and challenges related to AI integration, pointing out that the lack of infrastructure, the digital literacy skills gap and no clear ethical guidelines hindered the pedagogical contribution of such technologies at large.

As a whole, the reviewed studies in this article show that AI has been embraced in Saudi research writing, though not uniformly. Household names in text generation tools, like ChatGPT, Grammarly and QuillBot, are prevalent while more sophisticated academic-focused platforms like Scite are less common (Alghamdi & Abusa’Aleek, 2025; Fadlelmula & Qadhi, 2024). While these reflect general trends of adoption, they also highlight structural particularities related to Saudi context pressures (bilingual Arabic–English writing, infrastructural differences between universities and a regulatory regime still in process).

However, the Saudi literature on AI-assisted writing is scattered. Large-scale studies are limited, and most of the studies involve small samples and are cross-sectional in nature, recruiting undergraduate or ESL students (Ozfidan et al., 2024; Abduljawad 464, 2024). Although they grant an appreciation of short-term perceptions and performance, the latter measures don't correlate with such outcome variables as the quality of the thesis or the acceptance rate, nor long-term scholarly productivity. Moreover, methodological diversity is limited. Descriptive statistics or explorative factor analysis, rather than strong experimental or longitudinal designs, are used in many studies (Hamoda et al., 2025).

Supervisor perceptions, institutional policies, and ethical implications are particularly under-discussed dimensions. Alhusaiyan (2025) and Fadlelmula & Qadhi (2024) stress the lack of standardised disclosure guidelines, the ambiguous implementation of academic integrity policies, and the poor ability of teachers and students for training. These just-mentioned gaps evoke urgent issues: when do AI tools effectively improve the quality of research writing? How to address the risks of plagiarism and dependence? What institutional support and infrastructure does responsible AI need to become more embedded in academic work?

Furthermore, Saudi studies often focus on surface-level writing outcomes (grammar, fluency, time efficiency) while neglecting deeper aspects such as argumentation, methodological design, and originality (Abduljawad, 2024; Aljuaid, 2024). The result is a literature that highlights immediate utility but provides limited guidance for aligning AI use with Vision 2030’s strategic goals of research excellence and international competitiveness. Building on the research gaps identified, the present study pursues three interrelated objectives:

- To systematically identify and categorize the artificial intelligence tools most frequently used by Saudi researchers and students in supporting academic research writing, and to trace their distribution across the research-writing workflow (pre-writing, drafting, post-writing, and analytical stages).
- To synthesize empirical evidence on the documented benefits (e.g., improved linguistic quality, enhanced productivity, academic confidence) and challenges (e.g., plagiarism risks, reduced critical thinking, lack of institutional policy) associated with the adoption of AI tools in Saudi research writing.
- To highlight underexplored dimensions such as institutional policies, supervisor training, infrastructural readiness, and equity of access and to situate these within the broader Vision 2030 agenda, thereby generating actionable recommendations for policymakers, universities, and educators.

In response to these gaps, this systematic review seeks to consolidate and interpret the evidence from 19 empirical and review studies published between 2020 and 2025. The review is guided by two research questions:

RQ1: Which AI tools are most frequently used in Saudi academic research writing, and how are they distributed across different stages of the research-writing workflow?

RQ2: What are the documented benefits, challenges, and research gaps associated with the use of AI tools in supporting academic research writing in Saudi Arabia?

This review makes four primary contributions. First, it provides the first Saudi-centered synthesis of empirical research on AI-assisted academic research writing. While earlier reviews have focused on global contexts (Hamoda et al., 2025; Alanazi et al., 2025) or classroom writing support (Aljuaid, 2024), this study specifically examines research writing for scholarly dissemination journal articles, theses, dissertations, and grant proposals.

Second, it integrates both quantitative and qualitative findings. By synthesizing descriptive statistics, effect sizes, and qualitative themes, the review provides a nuanced understanding of how AI tools influence writing quality, productivity, and academic integrity. E.g., while ChatGPT and Grammarly were most frequently studied, their adoption clustered around drafting and post-writing stages, indicating an emphasis on surface corrections rather than end-to-end workflow integration.

Third, the review situates Saudi evidence within the framework of Vision 2030, emphasizing the alignment of AI adoption with national priorities for research excellence and digital transformation (Saudi Vision 2030; Ryan, 2023). This contextualization highlights leverage points for policymakers, such as investments in digital infrastructure, targeted training, and ethical guidelines.

Finally, the review generates practical recommendations for multiple stakeholders. For policymakers, it identifies gaps in disclosure policies and training frameworks. For universities, it underscores the need to balance access and equity in AI adoption. For supervisors, it highlights the importance of guiding students to use AI critically rather than dependently.

Definitions and Scope

For clarity, four definitional elements guide this review.

Academic research writing refers to scholarly texts intended for dissemination, including journal articles, theses, dissertations, conference papers, and grant proposals. Unlike classroom essays, these texts demand methodological rigor, disciplinary conventions, and originality (Aljuaid, 2024; Hamoda et al., 2025).

AI tools include applications that use NLP, machine learning, or algorithmic models to support writing tasks. These range from grammar checkers (Grammarly, Trinka) and paraphrasers (QuillBot, Wordtune) to LLM-based assistants (ChatGPT), literature discovery systems (Elicit, Scite, Consensus), and integrity tools (Turnitin, Criterion, Pigai) (Ozfidan et al., 2024; Alanazi et al., 2025).

Temporal scope covers studies published between January 2020 and January 2025. This period coincides with the release of transformer-based architectures and widespread adoption of GPT-3/4, marking a new era of AI integration into research writing. Pre-2020 literature is referenced only where relevant for historical context.

Geographic focus is restricted to Saudi Arabia, ensuring that findings and recommendations are locally relevant. Studies included in Table 2 feature Saudi students, faculty, or institutions explicitly, thereby capturing the unique opportunities and challenges of AI adoption in this context (Ozfidan et al., 2024; Abduljawad, 2024)

2. Literature Review

Global Landscape of AI in Academic Research Writing

The integration of AI tools for academic writing has increased dramatically over the past two decades, with a steepening inflexion following the release of transformer-based large language models (LLMs) like GPT-3 and GPT-4. These tools expanded the scope of academic writing support beyond surface-level grammar checkers and automated writing feedback (AWF) tools to act as a tool for ideation, drafting, editing and even synthesis with scholarly literature (Alanazi et al., 2025; Alharbi et al., 2024). International research evidence suggests that the use of tools such as Grammarly, Trinka and QuillBot in higher education and academic contexts results in improvements to grammar accuracy, text cohesion and writing fluency.

Meanwhile, systematic reviews suggest that these benefits in surface quality do not automatically equate to gains in deeper academic skills. AWF systems are sensitive to errors and well-suited to detect them; however, they appear incapable of promoting argumentation or originality (Alanazi et al, 2024). Similarly, Alharbi et al. (2024), in a pan-journal study of AI-based editorial support tools, note that only very few references are being made to ethical consequences or borders of authorship and misuse. This dialectic of technical assistance and epistemic giving establishes the tone for a great deal of international conversation.

Furthermore, due to other tools like Elicit, Scite, or Consensus, it is conceivable outside the boundaries of language. To discover literature and evidence. Globally, the literature raises one methodological concern; there are limited longitudinal studies (Hamoda et al., 2025) to confirm the end-on effect of AI- long-term impact on quality, acceptability, and portfolio output for research/ productivity.

In the meantime, issues of over-dependency, plagiarism, and ethical uncertainty are the talk of the academy. Elkhatah et al. (2023) cautioned that even AI text might not be detected by the classic measures for plagiarism, and Perkins (2023) stressed how programmes like ChatGPT can produce text that has the potential to make authorship assignment more complex. These challenges are further exacerbated by uneven institutional policies and insufficient attention to responsible AI use amongst academics and students alike (Hasan & Chocarro, 2024; Wang et al., 2024). So, while there is overwhelming global evidence of the potential AI holds in education, we are also reminded to be discerning and contextual in adopting rather than wholesale.

Typology of AI Tools in Academic Writing

The AI ecosystem for academic writing can be classified into four overlapping domains:

(a) Language-level tools. Grammarly, Trinko, and QuillBot provide grammar correction, style refinement, and paraphrasing. Google Translate remains widely used by bilingual and multilingual students to bridge linguistic gaps. Saudi studies mirror this global reliance: for instance, Abduljawad (2024) found that students using Grammarly and ChatGPT for revision achieved statistically significant gains in fluency and accuracy, while Mugableh (2024) reported enhanced vocabulary diversity when ChatGPT was introduced into EFL writing labs.

(b) Generative LLMs. ChatGPT, Copilot, and Writesonic exemplify tools capable of producing coherent drafts, rephrasing complex passages, or generating outlines. In Saudi contexts, ChatGPT emerged as the dominant tool, cited in more than half of the reviewed studies. Mugableh (2024) noted postgraduate students' reliance on ChatGPT for generating outlines, faster drafting and paraphrasing, while supervisors voiced concerns over disclosure practices.

(c) Literature discovery and synthesis tools. Applications such as Elicit, Scite, Consensus ai employ AI to assist in scoping reviews and evidence mapping. While their global relevance is rising (Hamoda et al., 2025).

(d) Integrity, referencing, and assessment tools. Turnitin, plagiarism detectors, and AI-detection platforms are primarily employed for academic integrity monitoring.

At the same time, this typology illustrates that while Saudi research engages across all four domains, tool adoption is uneven, with dominance of language-level and generative tools.

The Saudi Context: Patterns of AI Tool Use

Saudi Arabia's adoption of AI tools for academic writing is shaped by the broader Vision 2030 agenda, which prioritizes digital transformation and international competitiveness (Saudi Vision 2030; Ryan, 2023). The reviewed studies reveal patterns that both mirror and diverge from global trends.

The distribution of AI usage across research-writing stages shows a strong concentration at the drafting and post-writing levels. More than half of Saudi studies report use of AI for text generation, paraphrasing, and stylistic improvement (e.g., ChatGPT, Grammarly, QuillBot), while about 50% highlight revision, editing, and plagiarism detection (e.g., Grammarly, Turnitin) (Ozfidan & El-Dakhs, 2024; Al-Sofi, 2024). In contrast, only about a quarter of studies address pre-writing stages like idea generation and literature discovery, despite the documented value of Elicit, Scite, Consensus in accelerating evidence synthesis.

This imbalance underscores a pragmatic, utilitarian adoption of AI in Saudi contexts. Faculty and students employ AI primarily as a linguistic aid, addressing challenges of English as a Foreign Language (EFL), rather than as a means of enhancing methodological rigor or originality. Aljuaid (2024) explicitly cautioned against treating AI tools as substitutes for guided instruction, emphasizing the need to preserve critical thinking.

Reported Benefits in the Saudi Literature

A consistent finding across the 19 Saudi studies is that AI tools provide tangible benefits for students and researchers, particularly in the areas of language improvement, productivity, confidence, and scholarly support. The most prominent theme is enhanced grammar, fluency, and coherence. E.g., Abduljawad (2024) reported statistically significant writing quality improvements when students used Grammarly in combination with ChatGPT for revision activities. Al-Sofi (2014) also claimed that postgraduate students made significant vocabulary gains and compelling coherence level after they regularly used Grammarly and QuillBot during their classes. These results are consistent with the worldwide agreement on the utility of AI tools as linguistic scaffolding for non-native English writers.

And when it comes to being productive, AI tools also have a way of helping out. For example, ChatGPT and QuillBot are pitched as writer's block busters that can accelerate the writing process. Mugableh (2024) noted that Saudi writing lab students completed drafts faster. This work shows that AI supports the generation or serving of related text but also it reshapes students' academic relationship with research tasks and time so they might more readily move from idea to text.

Another popular benefit is higher confidence and involvement. The research has found that AI-supported drafting positively affects ESL learners' sense of self-efficacy in completing higher level writing tasks. e.g., Alghamdi & AbuSa'aleek (2025) found postgraduate students perceived to have improved on organization skills, when they had the help of ChatGPT; Alshehri (2025) proved that AI tools can be considered on improving rate and self-efficacy of students. This psychological dimension highlights the motivational implications of AI over a less threatening environment in writing. In addition, AI has been shown to increase the accuracy of references. Waghid (2024) suggested AI to liberate supervisory labour for conceptual and methodological innovation.

Documented Risks and Challenges

Despite these benefits, the Saudi literature consistently points to significant risks and challenges. Chief among these are plagiarism and originality issues. Rammuny and Handono (2023) and Fadlelmula and Qadhi (2024) have identified increased risks of plagiarism with the use of paraphrasing tools such as QuillBot and generative models like ChatGPT. Supervisors expressed unease that indistinct authorship lines and AI overuse could jeopardise the integrity of research (Polok & Dussin, 2025).

There is also a grave deterioration in critical thinking. Ozfidan and El-Dakhs (2024) discovered that students who overused AI tools in their project work also produced fewer deep arguments and engagement with evidence. (2024) also claimed that, despite the fact that ChatGPT had a stronger surface-level vocabulary, it did not have much impact on nurture personality of analytical rigour. These findings suggest that casual use of AI can simply delete the intellectual labor at the heart of research writing.

There is, too, the age-old issue of moral ambiguity. Aljuaid (2024) identified different levels of transparency on communicating the use of AI: Alharbi et al. (2024) who mentioned the global recommendations for authorship and integrity were not explicit. The supervisors working in Saudi Arabia also emphasize that specific policies should be set up to ensure transparency and ethical adherence when applied the ATWSL.

There are also institutional-level challenges and deficiencies in infrastructure. Alhusaiyan (2025) recorded a wide range in digital readiness by Saudi universities, and smaller colleges could not get easy access to the latest AI elements. Technical constraints such as low bandwidth are also hindering supportive adoption, and the difficulty in processing bilingual Arabic-English texts is another barrier. In support of this, Alyami et al (2025) included that there are large differences in digital literacy across students, leading to increased biases in who benefits the most from AI. It further exacerbates the imbalances that there are no systematized training options for students, as well as instructors.

Research Gaps and Underexplored Dimensions

Although AI tools are becoming increasingly popular, there are several literature gaps in Saudi region. For one, no longitudinal information is available. Most existing research are based on cross-sectional surveys (e.g., Ozfidan et al., 2024) where perceptions lasting are only captured at a snapshot in time, and do not examine longer-term consequences such as thesis quality, acceptance rates of publications etc or even the later - stage research productivity.

A second gap is the underrepresentation of postgraduate contexts. Much of the literature is concentrated on undergraduate English-as-a-Foreign-Language (EFL) learners, with limited attention to master's or doctoral-level research writing where the stakes for originality, rigor, and dissemination are significantly higher.

Third, there is limited methodological integration of AI into the research process. Few studies explore AI's potential role in research design, evidence synthesis, or data analysis, despite international examples such as pilot projects involving Elicit. This suggests a missed opportunity to investigate how AI could contribute to higher-order scholarly processes beyond text generation.

The fourth gap lies in faculty and supervisor perspectives. While select studies document faculty concerns about plagiarism and authenticity, systematic investigations into supervisory strategies and their adaptation to AI-supported environments remain rare.

Finally, policy frameworks are still evolving. Institutional responses vary widely, with universities experimenting with independent guidelines (McDonald et al., 2024). However, no national-level policies have been formalized, even though Vision 2030 places digital governance at the heart of educational transformation. Without standardized disclosure requirements and clear supervisory practices, AI adoption risks remaining fragmented and inconsistent across the Saudi higher education system.

Thematic Synthesis

The literature converges on several themes. First, Saudi adoption reflects a pragmatic orientation, with AI employed primarily for surface-level support. Second, challenges are systemic as well as pedagogical, encompassing not only students' overreliance but also institutional unpreparedness and infrastructural inequality. Third, there is a misalignment between practice and policy ambitions. Vision 2030 emphasizes innovation and global competitiveness, yet AI integration remains focused on remedial language support rather than transformative research practices.

Taken together, Saudi scholarship affirms international trends and shows local specificities grounded in EFL settings, institutional preparedness, and national policy emphasis. Addressing the identified gaps, especially through longitudinally driven, postgraduate-oriented research with policy knowledge, will be of significance in leveraging AI for support to academic writing activities among research writers in the Kingdom. Aggressive interventions and strong mechanisms need to be put in place to reverse this negative trend.

3. Methodology

The research used a protocol-led, systematic review approach supported by the PRISMA 2020 to ensure transparency, rigour, and replicability (Page et al., 2021). The review followed a desktop research approach, which emphasized the finding and synthesis of secondary quantitative data through peer-reviewed sources rather than direct collection (see Patra et al., 2007). Desktop research is particularly suitable for new and emerging fields, as it is cost-effective, timely, and can uncover large-scale trends in a diverse literature (Wahid et al., 2023; Bearman et al., 2023; Kuleto et al., 2021; Ouyang et al., 2022). Under this concept, a structured search strategy was established for studies regarding the use of AI tools in academic research writing support from Saudi Arabia.

Sources of Data

Various academic databases were searched to maximize comprehensiveness. These were Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, Wiley Online Library, Taylor & Francis Online, ERIC, Google Scholar, and Semantic Scholar. Second, repositories like ResearchGate and CiteSeerX were explored to identify additional peer-reviewed studies. The inclusion of multiple databases broadened coverage and reduced the risk of publication bias, ensuring that both global studies with a Saudi context and region-specific investigations were included.

Search Strategy and Criteria

The search strategy combined Boolean operators with targeted keywords, including “AI tools in academic writing,” “ChatGPT Saudi Arabia,” “Grammarly and student writing,” “QuillBot in academic research writing,” “AI tools and EFL learners Saudi Arabia,” “AI-assisted academic publishing,” and “ethical issues of AI in academic writing.” This strategy ensured the retrieval of a wide but relevant pool of studies, covering both pedagogical and ethical dimensions of AI-assisted writing practices in Saudi universities.

The inclusion criteria required that studies be published between January 2020 and January 2025, written in English, peer-reviewed, and focused explicitly on AI tools in supporting academic research writing within Saudi Arabia. This time frame was chosen to capture the period of accelerated growth of transformer-based AI tools such as GPT-3 and GPT-4 (Alharbi et al., 2024; Alanazi et al., 2025). Excluded materials comprised grey literature such as blogs, magazines, and opinion pieces, as well as non-peer-reviewed sources, pre-2020 publications, and works outside the academic context. Table 1 summarizes the inclusion and exclusion parameters that structured the selection process.

Table 1. Inclusion and Exclusion Criteria for Article Selection

Criterion	Inclusion	Exclusion
Topic	Studies exploring AI tools in academic writing and research support in Saudi Arabia (universities and academic institutions)	AI use in non-academic, commercial, or general adoption contexts not related to research writing
Publication	Peer-reviewed journal articles and conference proceedings	Books, blogs, magazines, grey literature, and non-peer-reviewed sources
Years	Published between January 2020 – January 2025	Studies published before 2020
Language	English-language publications	Non-English studies
Sector	Saudi universities and academic institutions	Private companies, schools (K–12), or non-academic settings

Quality Appraisal

All studies included, post first phase of screening, were quality appraised in order to ensure methodological rigour and context relevance using a structured rating system. The identification, screening, eligibility and inclusion process was conducted according to the PRISMA 2020 framework (Page et al., 2021), with assessments made on individual study quality using Critical Appraisal Skills Programme (CASP) checklists. The CASP instrument concentrated on key elements including being clear about the aims of the research, the appropriateness of the research design, adequacy of recruitment to participants, consideration of ethical issues and if there was consistency between data and findings.

At this stage, studies that did not report the methods clearly, described only were provided with or no empirical data. As a consequence of this review, the final evidence base consisted only of papers that satisfied the minimal quality criteria and directly focused on the utility of AI tools in scientific writing within Saudi Arabia. This title results from this work, thus the synthesis is based on a strong and situation specific ground and can support strong conclusions.

Data Analysis

The data were analysed using a structured document analysis, which allows the systematic tracing of patterns and development of coherent themes in the results (Morais et al., 2021). A total of 19 peer-reviewed studies were included after the PRISMA 2020 protocol and CASP checklists for quality assessment. This restriction meant that the evidence base comprised only those articles which were methodologically rigorous and thematically pertinent.

The coding framework was developed based on the research questions. It means that every article was read, concepts were excerpted (coded), and a summary was made. Codes were developed and refined iteratively, and any redundant or irrelevant codes were eliminated. Four guiding paradigms were drawn: AI adoption within the ecology of a writing-research process; student and instructor stances toward AI integration; pedagogical stakes in tool use; ethics and institutional implications (including those for plagiarism as well as principles of authorship attribution and disclosure).

Quantitative and qualitative analyses were used to enhance the synthesis. Statistical findings. Descriptive statistics were used to determine the patterns of tool usage and thematic predominance among included studies. We then looked at how often different tools were used, and we also examined the frequency with which specific narratives or themes emerged. The coding framework was used thematically to capture recurring dimensions, including integration of AI into writing workflows, user perceptions, pedagogical issues, and ethical concerns.

Qualitatively, the examination found a range of context-specific practices among Saudi institutions. Some studies focused on increases in efficiency and enhanced fluency in writing processes, whereas others addressed potential threats: decreased cognitive work (neglect of revision and restructuring) and dependence on automated feedback. Through the integration of numeric trends and interpretive perspectives, this produced an evidence-based and situated synthesis. This double-layered tactic was done to ensure that the results were not only descriptive but also analytically sound for a full picture of how AI tools might be contributing to academic research writing in Saudi Arabia.

Research Procedures

According to the PRISMA 2020 guideline (Page et al., 2021), the review proceeded in four steps: identification, screening, eligibility, and inclusion. This mechanism guaranteed the resulting dataset to be methodologically strong and directly related to the research questions.

Identification: A comprehensive search was conducted across Scopus, Web of Science, IEEE Xplore, ERIC, SpringerLink, Taylor & Francis, Google Scholar, Semantic Scholar, and ResearchGate. Boolean operators and targeted keywords (e.g., “AI tools in academic writing,” “ChatGPT Saudi Arabia,” “Grammarly and student writing,” “ethical issues of AI in research writing”) were applied. This initial search yielded approximately 950 records.

Screening: Duplicate records were removed, leaving about 750 unique publications. Titles and abstracts were assessed against inclusion/exclusion criteria. Studies published before 2020, non-peer-reviewed materials (e.g., blogs, editorials, grey literature), and research outside the Saudi academic context were excluded.

Eligibility: A total of 120 full-text articles were examined independently by two reviewers. Disagreements were resolved by consensus. Excluded studies were largely those focusing on non-academic writing domains (e.g., journalism, creative writing) or K-12 education.

Inclusion: Ultimately, 19 studies met all criteria. These covered a range of tools: ChatGPT, Grammarly, QuillBot, Turnitin, and less frequently, Google Translate, Elicit, Scite, Consensus, or citation managers.

To ensure reliability, a standardized data extraction form captured bibliographic information, study design, participant demographics, tools investigated, and outcomes. All studies underwent CASP appraisal, which evaluated clarity of research questions, methodological appropriateness, ethical considerations, and validity of outcomes.

This rigorous multi-stage procedure, narrowing nearly one thousand records to 19 high-quality studies, ensured transparency, replicability, and international alignment. The final dataset provided a robust evidence base for analyzing how AI tools are supporting academic research writing practices in Saudi Arabia.

4. Findings

Study Selection

The identification and selection of studies followed the structured sequence of the PRISMA 2020 framework to ensure transparency and reproducibility (Page et al., 2021). A comprehensive database search across Web of Science, Scopus, ERIC, and Google Scholar initially produced 950 records, reflecting the rapid expansion of scholarship on AI-assisted academic writing and the broad relevance of AI tools across disciplines (Alanazi et al., 2025; Hamoda et al., 2025).

Automated filters and manual screening were used to identify and remove duplicates, eliminating 200 overlapping records and leaving 750 unique records for title–abstract screening. Using the a priori inclusion and exclusion criteria (Table 1), 630 items were excluded at this stage. Common reasons included publication prior to 2020 outside the period in which transformer-based systems (e.g., GPT-3/4) reshaped writing practices (Alharbi et al., 2024) a lack of direct focus on AI in

supporting academic research writing (e.g., studies on general ed-tech or non-writing applications), and materials classed as grey literature (e.g., blogs, editorials, magazines), which were excluded to preserve methodological rigor (Bearman et al., 2023).

At the eligibility stage, 120 full texts were retrieved and examined independently by two reviewers, with discrepancies resolved through consensus, in line with best practice for systematic reviews (Zhao et al., 2022). 65 studies were excluded at this point, primarily because they centered on non-academic writing contexts (e.g., journalism, SEO, creative writing) or were situated in corporate/K-12 settings rather than Saudi universities and academic institutions.

In accordance with the quality-appraisal refinement described in the Methodology (CASP checklists), a second-level appraisal was conducted after the PRISMA eligibility stage to enhance evidential robustness. This refinement excluded a further 36 studies (e.g., non-peer-reviewed outputs, insufficient methodological detail, or indirect relevance to AI-supported research writing in Saudi Arabia). The final dataset comprised 19 peer-reviewed studies, which form the evidence base for the synthesis presented in the remainder of the Results. The PRISMA flow diagram (Figure 1) summarizes the process from identification to final inclusion, including the post-eligibility quality-appraisal refinement.

Inclusion and exclusion criteria (Table 1) were applied consistently to all records, and the final set of 19 studies reflects the most methodologically sound and contextually relevant evidence on AI tools supporting academic research writing in Saudi Arabia. Table 2 maps the characteristics of the 19 included studies (designs, tools, and focal outcomes) and provides the foundation for the recalculated distributions presented in Table 3 (AI tools across studies) and Table 4 (stages of the research-writing workflow). Subsequent sections report impacts and challenges with updated counts ($n = 19$) and percentages.

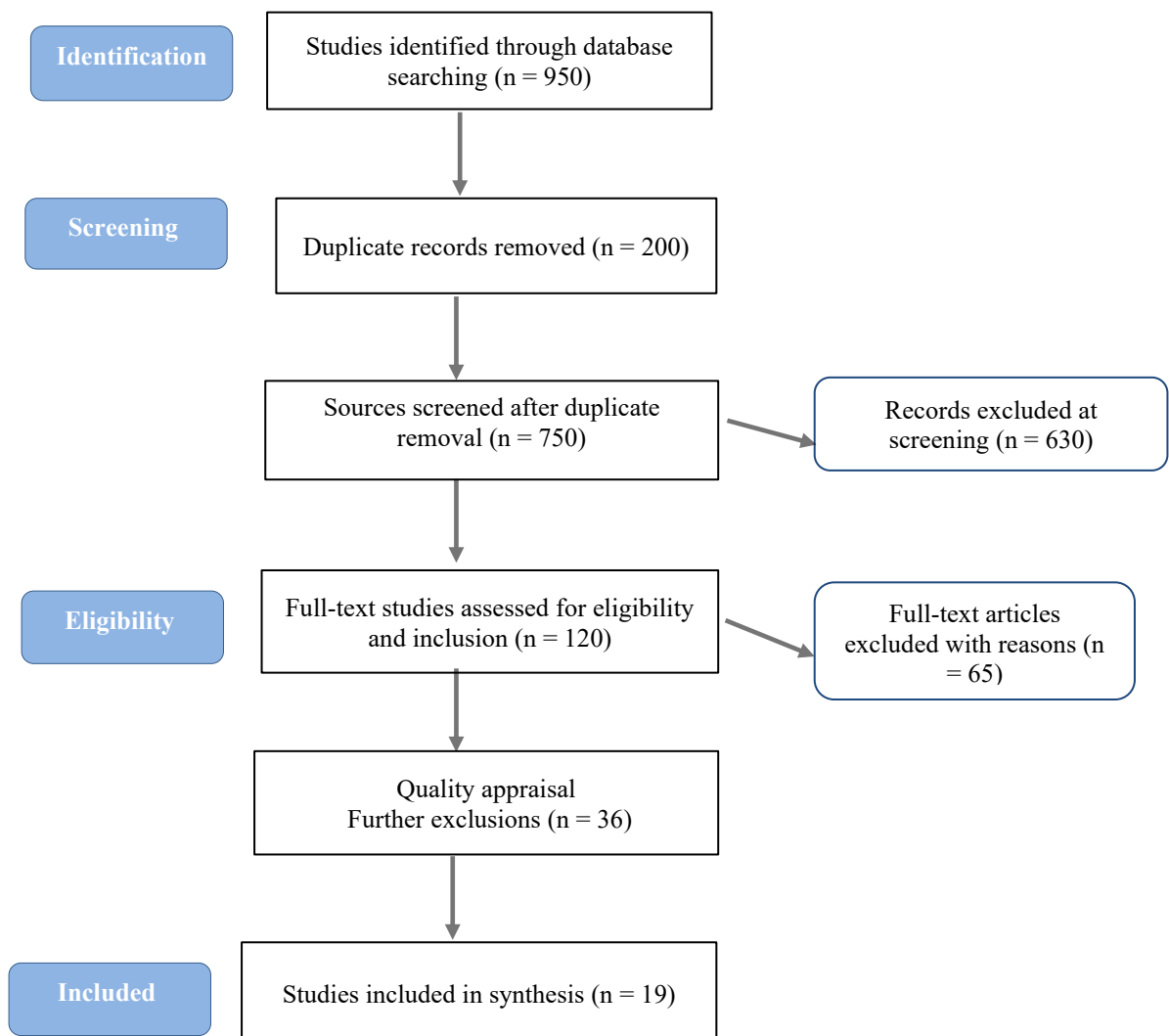


Figure 1. PRISMA 2020 flowchart of the article selection process

Study Characteristics

Table 2 provides an overview of the 19 studies that met the inclusion criteria, revealing clear methodological and topical trends. Most relied on cross-sectional surveys and descriptive evaluations, reflecting the exploratory stage of AI adoption in Saudi research-writing contexts (Hamoda et al., 2025). A smaller set used experimental or quasi-experimental designs to assess the impact of ChatGPT and Grammarly on fluency, coherence, and accuracy (Abduljawad, 2024; Masoudi, 2024). Only a handful incorporated mixed-methods or policy analyses, pointing to a gap in longitudinal and outcome-focused investigations (Alanazi et al., 2025). The majority of studies targeted undergraduate and postgraduate EFL/ESL learners, with limited attention to faculty practices, supervisory roles, or institutional policies, suggesting a research bias toward surface-level student support (Aljuaid, 2024). In terms of tools, ChatGPT and other LLMs dominate the literature, followed by Grammarly, QuillBot, and Turnitin, literature discovery platforms Elicit, Scite, Consensus. This imbalance highlights a prevailing focus on tools that improve surface accuracy and productivity, rather than those enhancing originality, evidence synthesis, or analytical rigor (Alanazi et al., 2025; Hamoda et al., 2025). This table represents the core evidence base of the review and directly reflects the methodological filters outlined in Table 1.

The majority of studies targeted undergraduate and postgraduate EFL/ESL learners, with limited attention to faculty practices, supervisory roles, or institutional policies suggesting a research bias toward surface-level student support (Aljuaid, 2024). In terms of AI tools, ChatGPT appeared in 8 of 19 studies (42.1%), followed by Grammarly (7; 36.8%), QuillBot (2; 10.5%), and Turnitin (2; 10.5%). Tools supporting literature discovery (Elicit, Scite, Consensus) appeared in only 1 study. This imbalance highlights a prevailing focus on tools that enhance surface accuracy and productivity, rather than those supporting originality, evidence synthesis, or analytical rigor (Alanazi et al., 2025; Hamoda et al., 2025). This table represents the core evidence base of the review and reflects the methodological filters outlined in Table 1.

Table 2. Study characteristics of the included studies (n = 19)

N.	Author(s), Year	Study Design	AI Tool(s)	Key Findings
1	Masoudi, 2024	Experimental (Saudi EFL learners)	ChatGPT	ChatGPT feedback significantly improved writing fluency compared with control groups.
2	Alharbi et al., 2024	Cross-sectional survey (healthcare students)	ChatGPT	Positive attitudes, but concerns about accuracy & ethics; highlighted barriers in Saudi context
3	Ozfidan, El-Dakhs & Alsalm, 2024	Survey (Saudi undergraduates)	ChatGPT, Grammarly, Google Translate	Two latent factors identified: instructional support and instructional practices.
4	Abduljawad, 2024	Experimental (Saudi ESL students)	Grammarly, ChatGPT	Significant gains in accuracy and fluency of student writing.
5	Alhabib & Alghammas, 2023	Experimental (Saudi undergraduates)	Grammarly	Automated feedback improved grammar & writing skills
6	Hamoda et al., 2025	Expert perspective	Elicit, Scite, Consensus	Literature-discovery tools beneficial but risk of uncritical automation.
7	Aljuaid, 2024	Review (Saudi focus)	AI writing assistants	Recommends AI as supplements, not substitutes, to preserve critical thinking.
8	Alhusaiyan, 2025	Policy/educational analysis	AI in writing broadly	Identified digital literacy gaps and ethical ambiguities in Saudi AI adoption.

9	Fadlelmula & Qadhi, 2024	Institutional policy study	AI tools in universities	Advocated training and ethical frameworks aligned with Vision 2030.
10	Al-Sofi, 2024	Mixed methods	Grammarly, QuillBot	Tools improved coherence and vocabulary among postgraduates.
11	Alanazi et al., 2025	Mixed-methods	ChatGPT (LLM-powered automated feedback)	Documented shift to LLM-powered feedback; implications for Saudi research.
12	Mugableh, 2024	Controlled pilot (writing lab)	ChatGPT	Faster drafting and vocabulary diversity; weak effect on argument depth.
13	Alghamdi & Abusa'Aleek, 2025	Graduate survey	Grammarly, Turnitin	Grammarly seen as editing aid; Turnitin valued mainly as deterrent.
14	Alshehri, 2025	Mixed-methods	Multiple AI-writing tools	Gains in speed and fluency; concerns about skill erosion.
15	Syed et al., 2024	Cross-sectional survey	ChatGPT	Tools aided proposal drafting; authors highlighted need for ethical and responsible AI use in research.
16	Alsofyani & Barzanji, 2025	Mixed methods	ChatGPT	ChatGPT feedback was as effective as teacher feedback; students reported positive perceptions and valued immediacy and personalization.
17	Alyami et al., 2025	Mixed methods, including questionnaires and semi-structured interviews	Grammarly	Grammarly improved procedural writing but needed human feedback.
18	Polok & Dussin, 2025	Comparative governance analysis	AI ethics and generative AI applications	Supervisors showed pragmatic acceptance, with calls for disclosure rules.
19	Alharbi, 2023	Pedagogical overview	Automated writing assistance tools (e.g., Grammarly, citation managers)	Highlighted potential in Saudi EFL classrooms to improve academic writing and referencing.

Based on the AI tools explicitly identified in the 19 studies included in this review (as outlined in Table 2), Table 3 demonstrates a clear distribution emerges in terms of tool popularity and application. ChatGPT is the most frequently used tool, cited in 8 studies (42.1%), highlighting its central role in drafting, paraphrasing, and automated feedback. Grammarly follows closely, mentioned in 7 studies (36.8%), and is mainly employed for grammar correction, coherence enhancement, and fluency improvement.

Other tools were mentioned far less frequently. Google Translate and Elicit were each cited in just one study (5.3%), suggesting minimal focus on pre-writing support or literature discovery tools. Meanwhile, a category labeled "Other AI Tools" (e.g., citation managers, policy-level AI systems, or unspecified assistants) was noted in 2 studies (10.5%), capturing cases where AI use was discussed broadly without naming specific platforms.

This distribution illustrates that most Saudi academic writing research focuses on tools that support surface-level language enhancement, with far fewer studies examining tools that aid in research design, synthesis, or analytical depth.

Table 3. Distribution of AI Tools Across Included Studies (n = 19)

AI Tool	No. of Studies	Percentage (%)
ChatGPT	8	42.1%
Grammarly	7	36.8%
Google Translate	1	5.3%
Elicit	1	5.3%
Other AI Tools	2	10.5%
Total	19	100.0%

Beyond identifying which AI tools are used, it is equally important to understand how these tools are applied across different stages of the academic research-writing workflow. Based on the analysis of the 19 included studies, Table 4 shows the drafting stage emerged as the most frequently supported, reported in 11 studies (57.9%). At this stage, tools such as ChatGPT and Grammarly were commonly used to assist with text generation, fluency, and paraphrasing.

Post-writing activities (revision, editing, and plagiarism checks) were supported in 8 studies (42.1%), with Grammarly and Turnitin being the most widely cited tools for surface-level corrections and integrity assurance. In contrast, pre-writing support such as ideation and literature discovery appeared in only 3 studies (15.8%), with tools like Elicit and Google Translate assisting early-stage writing tasks. Finally, analytical functions (e.g., coding, referencing, and synthesis) were mentioned in just 2 studies (10.5%), reflecting a continued underuse of AI in deeper research tasks such as qualitative analysis or citation management.

This distribution reveals a pragmatic focus on language enhancement and workflow efficiency, with relatively limited engagement in higher-order research competencies like argument development, evidence synthesis, or critical analysis.

Table 4. Stages of the Research-Writing Workflow Supported by AI Tools (n = 19)

Stage of Writing Workflow	Main AI Tools Identified	No. of Studies	Percentage (%)
Pre-writing (Ideation & Literature Discovery)	ChatGPT (idea generation), Elicit, Google Translate	3	15.8%
Drafting (Writing & Paraphrasing)	ChatGPT, Grammarly, QuillBot	11	57.9%
Post-writing (Revision & Integrity Checks)	Grammarly, Turnitin	8	42.1%
Analytical (Data Analysis & Referencing)	Citation managers, Scite	2	10.5%
Total	-	19	100.0%

Note: Percentages reflect the number of studies that explicitly addressed a stage; some studies covered multiple stages.

After identifying where AI tools are used and how they function across writing stages, Table 5 summarizes the positive and negative impacts reported across the 19 studies.

On the positive side, the most frequently cited benefit was improved language quality, such as better grammar, coherence, and fluency reported in 9 studies (47.4%) and primarily associated with Grammarly and ChatGPT. Increased productivity

and time efficiency appeared in 6 studies (31.6%), particularly in the use of ChatGPT for faster drafting and revision. Academic confidence and motivation were noted in 4 studies (21.1%), as students felt more capable and encouraged by real-time feedback and AI support.

Fewer studies emphasized deeper research benefits. Support in idea generation or literature discovery was mentioned in 2 studies (10.5%), typically involving ChatGPT or Elicit. Similarly, referencing accuracy and feedback quality (e.g., via Turnitin or citation tools) appeared in 2 studies (10.5%).

On the negative side, the most pressing concern was plagiarism and authorship ambiguity, reported in 6 studies (31.6%), particularly where students relied heavily on QuillBot or failed to disclose ChatGPT use. Reduced critical thinking appeared in 4 studies (21.1%), with concerns that AI outputs sometimes replaced genuine analytical work. Institutional and policy-related gaps were flagged in 3 studies (15.8%), reflecting the lack of formal AI-use guidelines. Digital literacy and infrastructure limitations (such as poor access to tools or bandwidth issues) were also noted in 3 studies (15.8%), especially in smaller institutions. Lastly, AI bias and output inaccuracies (e.g., hallucinations) were reported in 2 studies (10.5%), raising concerns over unchecked AI-generated content.

Overall, the findings suggest that while AI tools offer substantial linguistic and productivity benefits, their epistemic and ethical implications remain unresolved in the Saudi academic context.

Table 5. Reported Impacts of AI Tools on Academic Research Writing in Saudi Arabia (n = 19)

Impact Type	Key Outcomes Identified	No. of Studies	Percentage (%)
Positive Impacts			
Improved language quality (grammar, fluency, coherence)	Grammarly, ChatGPT	9	47.4%
Increased productivity & time efficiency	ChatGPT	6	31.6%
Enhanced academic confidence & engagement	ChatGPT, Grammarly	4	21.1%
Support in idea generation & literature discovery	ChatGPT, Elicit	2	10.5%
Strengthened feedback & referencing	Turnitin, citation tools	2	10.5%
Negative Impacts			
Risk to originality / plagiarism concerns	ChatGPT, QuillBot, Turnitin	6	31.6%
Lack of critical thinking / deep argumentation	Over-reliance on AI	4	21.1%
Institutional & policy gaps	Absence of disclosure rules	3	15.8%
Infrastructure & digital literacy issues	Bandwidth, access inequality	3	15.8%
Bias & accuracy limitations of AI tools	AI hallucinations, translation errors	2	10.5%

As shown in Table 6, the most frequently cited concern is related to ethical and academic integrity risks, highlighted in 6 studies (31.6%). These include issues of plagiarism, ambiguous authorship, and insufficient disclosure when AI-generated content is used raising questions about accountability and scholarly transparency (Aljuaid, 2024; Alanazi et al., 2025).

Pedagogical challenges were noted in 5 studies (26.3%), particularly regarding over-reliance on AI tools like ChatGPT or QuillBot, which may weaken students' originality, argumentation, and critical thinking skills. Meanwhile, research gaps in the literature were observed in 5 studies (26.3%) as well, with many studies lacking longitudinal insight into AI's effects on thesis quality, publication outcomes, or postgraduate-level writing practices.

Institutional and policy inconsistencies appeared in 4 studies (21.1%), reflecting a lack of unified guidelines on ethical AI usage across Saudi universities. Technical and infrastructure limitations, including unequal tool access, low bandwidth, and limited bilingual processing capabilities, were documented in 3 studies (15.8%). Furthermore, training and awareness gaps (noted in 3 studies) continue to hinder responsible and critical use of AI tools by both students and supervisors. Finally,

equity and access issues, such as the digital divide between well-resourced and under-resourced institutions, were highlighted in 2 studies (10.5%).

These findings confirm that while AI tools offer valuable benefits, the sustainable and ethical integration of AI into research writing in Saudi Arabia requires systemic interventions particularly in policy development, faculty training, and equitable digital infrastructure.

Table 6. Challenges and Research Gaps Identified in Saudi Studies (n = 19)

Category	Specific Challenges / Gaps	No. of Studies	Percentage (%)
Ethical & Integrity Concerns	Plagiarism risks, unclear authorship, low disclosure, AI hallucinations	6	31.6%
Pedagogical & Skills	Over-reliance on AI reduces critical thinking, originality, and academic writing skills	5	26.3%
Research Gaps in Literature	Lack of longitudinal studies, limited evidence on postgraduate writing and publication impact	5	26.3%
Institutional & Policy Gaps	Absence of clear guidelines; inconsistent enforcement across universities	4	21.1%
Technical & Infrastructure	Limited access to AI tools; weak digital infrastructure; bilingual text limitations	3	15.8%
Training & Awareness Deficits	Lack of structured AI training for students and supervisors	3	15.8%
Equity & Access Issues	Disparities between large and small universities; unequal AI literacy	2	10.5%
Total	-	19	100%

5. Discussion

This systematic review explored the integration of artificial intelligence (AI) tools in supporting academic research writing across Saudi Arabia. Synthesizing evidence from 19 included studies, it examined (1) which tools are most commonly used and at what stages of the research-writing workflow (RQ1), and (2) what benefits, challenges, and gaps emerge from their application (RQ2). The findings reveal a pragmatic adoption landscape, where tools are used primarily to enhance linguistic quality and streamline productivity, while deeper functions such as research synthesis, analytical support, and critical reasoning remain underutilized. These trends mirror broader dynamics in AI adoption globally but are shaped by local pedagogical, linguistic, and infrastructural factors.

RQ1: Which AI Tools Are Used and Where in the Workflow?

The tools most frequently cited in the Saudi literature are those that offer immediate and surface-level improvements to writing, particularly ChatGPT, Grammarly, and QuillBot. As shown in Table 3, ChatGPT appeared in 8 out of 19 studies, serving as a generative writing assistant for brainstorming, outlining, and drafting (Masoudi, 2024; Mugableh, 2024). Grammarly, reported in 7 studies, was the most popular tool for grammar correction, coherence refinement, and style consistency (Abduljawad, 2024; Al-Sofi, 2024). QuillBot, while less frequently cited, was praised for its paraphrasing capabilities, particularly in postgraduate contexts (Alghamdi & Abusa'Aleek, 2025).

The usage of these tools maps clearly onto the research-writing workflow. As outlined in Table 4, the drafting stage was the most heavily supported (11 studies, 57.9%), with students and researchers using generative AI to overcome writer's block, rephrase text, or accelerate content development. Post-writing activities such as editing, proofreading, and plagiarism checks were also common (8 studies, 42.1%), relying on tools like Turnitin and Grammarly to refine final drafts and ensure academic integrity (Aljuaid, 2024; Alghamdi & Abusa'Aleek, 2025).

By contrast, pre-writing support, including ideation, topic exploration, and literature discovery, was explicitly addressed in only three studies (15.8%), despite the availability of tools such as Elicit and Google Translate. Likewise, analytical and

referencing functions were identified in just two studies (10.5%) typically via citation managers or Scite, signaling minimal engagement with AI for coding, synthesis, or citation classification. (Hamoda et al., 2025). Collectively, these patterns suggest that AI is still treated primarily as a writing assistant rather than a cognitive collaborator across the broader research process (Aljuaid, 2024).

Several contextual factors explain this pattern. First, the bilingual demands of academic publishing in English, often faced by Saudi EFL learners, increase pressure on surface accuracy and fluency (Al-Sofi, 2024). Second, time constraints and productivity pressures especially for postgraduate students encourage reliance on tools that yield immediate benefits. Third, digital literacy and training gaps limit the effective use of advanced tools, especially in stages requiring judgment or interpretive skill (Alshehri, 2025).

Overall, the workflow landscape in Saudi academic contexts reflects a functional concentration on drafting, editing, and submission integrity, with relative neglect of pre-writing ideation and analytical depth. This pattern, while pragmatic, limits the transformative potential of AI to enhance research thinking and methodological rigor (Alanazi et al., 2025).

RQ2: What Are the Benefits, Challenges, and Gaps in AI Use?

The review identified a range of positive impacts resulting from AI use, particularly in improving language quality, productivity, and writer confidence. As detailed in Table 5, the most frequently reported outcome was enhanced grammar, coherence, and fluency, supported by Grammarly and ChatGPT, and cited in 9 studies (47.4%) (Abduljawad, 2024; Al-Sofi, 2024). Several studies also documented improved productivity and time efficiency (6 studies, 31.6%), with students noting faster drafting, reduced anxiety, and easier revisions (Mugableh, 2024; Alshehri, 2025).

Additionally, academic confidence and motivation were observed to rise when students used AI tools for outlining or early drafting stages. The perception of increased fluency and speed translated into higher engagement levels, particularly among postgraduate students (Alghamdi & Abusa'Aleek, 2025; Syed et al., 2024). Faculty members also highlighted the value of delegating low-level corrections to AI, thereby freeing time for deeper supervisory feedback on structure, argumentation, and methodology (Makarius et al., 2020).

Despite these benefits, the literature also presents a set of persistent challenges. The most serious of these relate to academic integrity. In 6 studies (31.6%), researchers expressed concern over unclear authorship, paraphrasing misuse, and inadequate disclosure of AI-assisted work (Fadlelmula & Qadhi, 2024; Alharbi et al., 2024). These issues are exacerbated by inconsistent institutional policies, which leave both students and faculty uncertain about ethical boundaries (Aljuaid, 2024; McDonald et al., 2024).

Another critical issue is the erosion of critical thinking and original argumentation. While AI tools improve surface-level quality, their heavy use is associated with shallow analytical reasoning and reliance on generic content. In classroom and lab settings, this was evident in the contrast between improved fluency and stagnant reasoning complexity (Ozfidan & El-Dakhs, 2024; Mugableh, 2024).

Technical and access challenges further compound these risks. As detailed in Table 6, infrastructure limitations, particularly in smaller or under-resourced institutions, limit access to advanced tools and stable internet connectivity (Alhusaiyan, 2025). Simultaneously, disparities in digital literacy mean that tech-savvy students benefit disproportionately from AI, creating new equity gaps in writing performance (Alyami et al., 2025).

From a research standpoint, the review highlights several critical gaps. First, there is a lack of longitudinal studies; most current research uses cross-sectional designs and self-report surveys that fail to capture long-term effects on thesis quality, publication success, or scholarly development (Ozfidan et al., 2024). Second, supervisory and postgraduate contexts are underexplored. While undergraduate EFL populations dominate the evidence base, little is known about how supervisors navigate AI use in dissertations or journal writing (Polok & Dussin, 2025).

Third, methodological integration of AI tools remains limited. Tools like Elicit and Scite, which could enhance evidence synthesis, qualitative coding, and citation management, appear only in pilot projects or expert commentaries (Hamoda et al., 2025). Fourth, there is a striking absence of policy evaluation, despite the growing urgency of standardizing AI disclosure and authorship norms across Saudi universities (McDonald et al., 2024; Alhusaiyan, 2025).

These gaps correlate with the neglected workflow stages outlined under RQ1 particularly pre-writing and analysis—and help explain why studies consistently report strong gains in grammar and efficiency, but weaker improvements in originality, argumentation, or epistemic rigor.

Synthesis and Implications

Taken together, the Saudi evidence base portrays AI as a supplement—not a substitute for academic writing. When embedded within structured guidance, AI tools such as Grammarly and ChatGPT provide effective support for revision and clarity (Alshehri, 2025; Al-Sofi, 2024). However, when used uncritically or excessively, especially for paraphrasing or full drafting, AI risks displacing student thinking and compromising authorship (Fadlelmula & Qadhi, 2024; Ozfidan & El-Dakhs, 2024).

The value of AI thus hinges not on the tool itself, but on the conditions of use including user training, institutional policies, and supervisory mediation. Without these, AI may accelerate “thin writing” text that is fluent but shallow in reasoning or originality.

Based on the reviewed evidence, three key practice-oriented implications emerge for enhancing the integration of AI in supporting academic research writing within Saudi Arabia. First, there is a clear need to expand AI integration across the entire research-writing workflow. Current usage is heavily concentrated in drafting and editing stages, but greater benefits may be unlocked by incorporating AI tools during pre-writing (e.g., Elicit, Scite, Consensus) and analytical phases (e.g., NVivo, ATLAS.ti and citation managers). These tools can support ideation, literature mapping, and data analysis, which are currently underutilized areas. However, such expansion requires not only improved access to these platforms but also targeted training to develop users’ interpretive and methodological skills, ensuring that AI outputs are used critically and appropriately.

Second, there is an urgent need to standardize policy and ethical guidance across institutions. The review exposes inconsistent disclosure efforts and obscure authorship standards that threaten academic ethics and test comprehension. The foundation of clear and actionable institutional (McDonald, 2024) policies for science writing assisted with AI should specifically include: guidelines related to authorship attribution, disclosure statements, and evaluation of performance. The third, equally crucial one is that the capacity and equity gaps must be fixed. The gap in both digital infrastructure and the literacy of AI magnifies to disproportionate benefits, # putting students from poorly resources schools at further disadvantage. "Infrastructure development must also be accompanied by AI literacy for students and supervisors to avoid increasing academic segregation (Alhusaiyan, 2025; Alyami et al., 2025).

6. Conclusion

This systematic review examined the integration of AI tools in supporting academic research writing in Saudi Arabia. The results indicate that AI solutions are being utilized more and more as supplementary instruments aimed at overcoming certain language and productivity obstacles instead of being seen as a replacement for the original academic writing tasks. For RQ1, we found that the top used tools, ChatGPT, Grammarly, and QuillBot, are also around drafting and post-writing stages of writing, supporting paraphrasing, grammar correction, and plagiarism checks. In contrast, tools for pre-writing discovery (e.g., Elicit) and analysis (e.g., Citation Managers, Scite, Consensus, and Elicit) continue to be lightly adopted – indicative of a limited functional range that is more focused on surface-level improvement rather than deep scholarly engagement.

Addressing RQ2, multiple studies reported improvements in language quality, efficiency, and student confidence as key benefits. However, significant challenges were also identified: concerns over academic integrity, lack of critical thinking, inconsistent disclosure policies, and unequal access to AI tools. These issues reflect broader structural gaps in digital infrastructure, policy clarity, and training preparedness, particularly for postgraduate and under-resourced institutions.

Overall, AI tools in Saudi academic writing function most effectively when treated as supplements under supervision, rather than autonomous substitutes. Their potential lies in lifting the technical baseline of writing while preserving scholarly integrity and critical reasoning. For sustainable integration, stakeholders must prioritize workflow-wide implementation, standardized ethical frameworks, and capacity-building initiatives. A balanced, policy-aligned model of AI adoption embedded in training, supervision, and equitable infrastructure can best support Saudi Arabia’s Vision 2030 and foster a more innovative, responsible research culture.

Reference:

- Abduljawad, S. A. (2024). Investigating the impact of ChatGPT as an AI tool on ESL writing: Prospects and challenges in Saudi Arabian higher education. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1–19. <https://www.researchgate.net/publication/387918568>
- Alanazi, S. M., Elmotri, B., Khamis, G. S., & Darem, A. A. (2025). Assessing the efficacy of ChatGPT's automated corrective feedback in enhancing students' writing proficiency. *International Journal of Advanced and Applied Sciences*, 12(2), 205–214. <https://doi.org/10.21833/ijaas.2025.02.022>
- Alghamdi, A. M., & AbuSa'aleek, A. O. (2025). Exploring postgraduates' attitudes towards Grammarly as an AWE tool in academic writing. *Studies in English Language and Education*, 12(2), 878–895. <https://doi.org/10.24815/siele.v12i2.39995>
- Alhabib, A., & Alghammas, A. (2023). Automated corrective feedback program (Grammarly): Effects on Saudi EFL undergraduates. *Migration Letters*. <https://doi.org/10.59670/ml.v20iS9.4823>
- Alharbi, M. (2024). The role of artificial intelligence in advancing English as a foreign language teaching at Saudi universities. *World Journal on Educational Technology: Current Issues*, 16(3), 181–200. <https://doi.org/10.18844/wjet.v16i3.9311>
- Alharbi, M. K., Syed, W., Innab, A., & Al-Rawi, M. (2024). Healthcare students' attitudes, opinions, perceptions and perceived obstacles regarding ChatGPT in Saudi Arabia: A survey-based cross-sectional study. *Scientific Reports*. <https://www.researchgate.net/publication/384536978>
- Alhusaiyan, E. (2025). A systematic review of current trends in artificial intelligence in foreign language learning. *Saudi Journal of Language Studies*, 5(1), 1–15. <https://doi.org/10.1108/SJLS-07-2024-0039>
- Aljuaid, H. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal, Special Issue on ChatGPT (April 2024)*, 26–55. <https://dx.doi.org/10.24093/awej/ChatGPT.2>
- Alshehri, M. (2025, March). Integrating Automated Writing Evaluation into Saudi EFL Students' Writing Practice. *Arab World English Journal*, 16(1), 70–87. <https://doi.org/10.24093/awej/vol16no1.5>
- Al-Sofi, B. B. M. A. (2024). Artificial intelligence-powered tools and academic writing: To use or not to use ChatGPT. *Saudi Journal of Language Studies*, 4(3), 145–161. <https://doi.org/10.1108/SJLS-06-2024-0029>
- Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*, 63(2), 431–463. <https://journals.sagepub.com/doi/abs/10.1177/07356331241307297>
- Alyami, A., Alotaibi, S., & Khan, W. (2025). Saudi EFL learners' perceptions of using AI and its impact on their writing skills. *Arab World English Journal*, 16(1), 349–365. <https://dx.doi.org/10.24093/awej/vol16no1.22>
- Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial Intelligence in higher education: a critical literature review. *Higher Education*, 86(2), 369–385. <https://doi.org/10.1007/s10734-022-00937-2>
- Cheong, C. M., Luo, N., Zhu, X., Lu, Q., & Wei, W. (2022). Self-assessment complements peer assessment for undergraduate students in an academic writing task. *Assessment & Evaluation in Higher Education*, 47(4), 642–656. <https://doi.org/10.1080/02602938.2022.2069225>
- Cumpston, M., McKenzie, J., Welch, V., & Brennan, S. (2022). Strengthening systematic reviews in public health: guidance in the Cochrane Handbook for Systematic Reviews of Interventions, 2nd edition. *Journal of Public Health (Oxford, England)*, 44(3), 588–592. <https://doi.org/10.1093/pubmed/fdac036>
- Deep, P. D., & Chen, Y. (2025). The Role of AI in Academic Writing: Impacts on Writing Skills, Critical Thinking, and Integrity in Higher Education. *Societies*, 15(9), 247. <https://doi.org/10.3390/soc15090247>

- El-Garawany, M. S. M. (2024). The effects of a QuillBot-based intervention on English academic writing performance. *Language Teaching Research Quarterly*, 43, 167-189. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1457355.pdf>
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19, Article 17. <https://doi.org/10.1007/s40979-023-00140-5>
- Fadlelmula, F., & Qadhi, S. (2024). A systematic review of research on artificial intelligence in higher education: Practice, gaps, and future directions in the GCC. *Journal of University Teaching and Learning Practice*, 21(6), Article 806. <https://doi.org/10.53761/pswgbw82>
- Fan, N. (2023). Exploring the effects of automated written corrective feedback via Grammarly on EFL students' writing quality: A quasi-experimental study. *SAGE Open*, 13(3). <https://doi.org/10.1177/21582440231181296>
- Hamoda, T. A., Kandeel, A., Elamin, A. M., Ghonim, Y. A., Elkashef, W. F., Fathi, A. K., ... & Shaer, O. (2025). Artificial Intelligence in Scientific Writing: Balancing Innovation and Efficiency with Integrity: Perspectives and Position Statements of Global Andrology Forum Expert Group. *World Journal of Men's Health*, published online Mar 27, 2025. <https://doi.org/10.5534/wjmh.240007>
- Hasan, M. R., & Chocarro, R. (2024). Understanding AI chatbot adoption in education: PLS-SEM approach. *Computers & Education: Artificial Intelligence*, 5, 100207. <https://doi.org/10.1016/j.caeai.2024.100207>
- Kuleto, V. et al. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability*, 13(18), 104-124. <https://doi.org/10.3390/su131810424>
- Masoudi, H. A. (2024). Effectiveness of ChatGPT in improving English writing proficiency among non-native English speakers. *International Journal of Educational Sciences and Arts*, 3(4). <https://doi.org/10.59992/IJESA.2024.v3n4p2>
- McDonald, N., Johri, A., Ali, A., & Hingle Collier, A. (2025, March). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3, 100121. <https://doi.org/10.1016/j.chbah.2025.100121>
- Mohiuddin, K., Nasr, O. A., Miladi, M. N., Fatima, H., Shahwar, S., & Naveed, Q. N. (2023). Potentialities and priorities for higher educational development in Saudi Arabia for the next decade: Critical reflections of the vision 2030 framework. *Heliyon*, 9(5). [https://www.cell.com/heliyon/fulltext/S2405-8440\(23\)03575-2](https://www.cell.com/heliyon/fulltext/S2405-8440(23)03575-2)
- Morais, F. P. de F., dos Santos, P. H. F., & Cauduro, F. L. F. (2021). Addressing themes related to stomatherapy in undergraduate nursing education: Document analysis. *ESTIMA, Brazilian Journal of Enterostomal Therapy*. https://doi.org/10.30886/estima.v19.1028_in
- Mugableh, A. I. (2024). The impact of ChatGPT on the development of vocabulary knowledge of Saudi EFL students. *Arab World English Journal*, Special Issue on ChatGPT (April 2024), 265–279. <https://dx.doi.org/10.24093/awej/ChatGPT.18>
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher Education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27(6), 7893–7925. <https://doi.org/10.1007/s10639-022-10925-9>
- Ozfidan, B., El-Dakhs, D. A. S., & Alsalm, L. A. (2024). The use of AI tools in English academic writing by Saudi undergraduates. *Contemporary Educational Technology*, 16(4), Article ep527. <https://doi.org/10.30935/cedtech/15013>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, 372, n71. <https://doi.org/10.1136/bmj.n71>



- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2), Article 7. <https://doi.org/10.53761/1.20.02.07>
- Polok, B., & Dussin, M. (2025). AI Governance in Saudi Arabia: Cultural Values and Ethical AI Regulations in Comparative Perspective. *Yearbook of Islamic and Middle Eastern Law Online*, 1(aop), 1-29. <https://doi.org/10.1163/22112987-bja00004>
- Ryan, M. (2023). Higher Education in Saudi Arabia: Challenges, Opportunities, and Future Directions. *Research in Higher Education Journal*, 43. <https://eric.ed.gov/?id=EJ1382950>
- Saudi Vision 2030: Kingdom of Saudi Arabia [Official document]. https://www.vision2030.gov.sa/media/rc0b5oy1/saudi_vision203.pdf
- Syed, W., Bashatah, A., Alharbi, K., Bakarman, S. S., Asiri, S., & Alqahtani, N. (2024). Awareness and Perceptions of ChatGPT Among Academics and Research Professionals in Riyadh, Saudi Arabia: Implications for Responsible AI Use. *Medical science monitor : international medical journal of experimental and clinical research*, 30, e944993. <https://doi.org/10.12659/MSM.944993>
- Teng, M. F., & Wang, C. (2023). Assessing academic writing self-efficacy belief and writing performance in a foreign language context. *Foreign Language Annals*, 56(1), 144–169. <https://doi.org/10.1111/flan.12638>
- Waghid, F. (2024). The use of artificial intelligence to augment student supervision. In *Philosophical Adventures in African Higher Education* (pp. 14). Routledge. <https://doi.org/10.4324/9781003530091-3>
- Wahid, N., Amin, U., Khan, M. A., Siddique, N., & Warraich, N. F. (2023). Mapping the desktop research in Pakistan: a bibliometric analysis. *Global Knowledge Memory and Communication*. <https://doi.org/10.1108/gkmc-07-2022-0159>
- Wang, C., Xu, W., & Li, H. (2024). Exploring students' generative AI-assisted writing processes and perceptions. *Smart Learning Environments*, 11, Article 29. <https://doi.org/10.1186/s40561-024-00317-6>
- Zhao, B. et al. (2022). Development of a reporting guideline for systematic reviews of animal experiments in traditional Chinese medicine. *Journal of Evidence-Based Medicine*, 15, 152 - 167. <https://doi.org/10.1111/jebm.12480>

"مراجعة منهجية لأدوات الذكاء الاصطناعي الداعمة للكتابة الأكاديمية البحثية في المملكة العربية السعودية"

اعداد الباحثة:

لمياء حسين الزهراني

الملخص:

تستعرض هذه المراجعة المنهجية الأدبيات المحكمة حول دور أدوات الذكاء الاصطناعي في دعم الكتابة الأكاديمية البحثية في المملكة العربية السعودية، مع تقييم فوائدها وتحدياتها والفجوات البحثية المرتبطة بها. جرى البحث في قواعد بيانات كبرى (مثل Scopus، Web of Science، IEEE Xplore، ERIC، Google Scholar) باستخدام تركيبات بوليانية مرتبطة بالذكاء الاصطناعي والكتابة الأكاديمية والتعليم العالي السعودي. اتُبع إطار PRISMA 2020 في الفحص، وقُيِّمت الجودة المنهجية باستخدام قوائم التحقق CASP. من أصل 950 سجلاً، استوفت 19 دراسة معايير الإدراج. معظم الدراسات كانت مقطعية أو تجريبية، وركزت على طلبة البكالوريوس والدراسات العليا من متعلمي اللغة الإنجليزية كلغة أجنبية أو ثانية (EFL/ESL).

أكثر الأدوات استخداماً شملت ChatGPT، Grammarly، QuillBot، Turnitin، والتي دعمت بشكل رئيسي مهام ما بعد الكتابة (مثل: إعادة الصياغة، تصحيح القواعد، تعزيز الطلاقة، وفحص الانتحال). في المقابل، حصلت الأدوات الموجهة إلى مرحلة ما قبل الكتابة (مثل Elicit، Scite، Consensus) وأدوات إدارة المراجع (مثل برامج إدارة الاستشهادات) على اهتمام محدود نسبياً، مما يشير إلى نمط تبني سطحي.

تمثلت الفوائد المبلغ عنها في تحسين الدقة اللغوية، الكفاءة الزمنية، وزيادة ثقة الطلاب بأنفسهم. كما أشار المشرفون إلى أن إسناد التصحيحات الميكانيكية إلى أدوات الذكاء الاصطناعي أتاح وقتاً أكبر لتقديم تغذية راجعة عالية المستوى بشأن البنية والحجج والمنهجية. ومع ذلك، برزت مخاطر متكررة، مثل: الانتحال وغموض مسألة التأليف، ضعف التفكير النقدي الناتج عن الإفراط في الاعتماد على المخرجات الآلية، تباين سياسات الإفصاح، إضافة إلى الفجوات في البنية التحتية ومهارات الثقافة الرقمية التي تحد من فرص التبني العادل.

بصورة عامة، تُظهر الأدلة أن أدوات الذكاء الاصطناعي تؤدي أفضل دور عندما تُستخدم كمكملات لا كبدايل للكتابة الأكاديمية. إن نموذجاً متوازناً متسقاً مع السياسات، يركز على التدريب، والوساطة الإشرافية، والإفصاح الشفاف، وضمان الوصول العادل، يعدّ ضرورياً لتحقيق دمج مستدام يتماشى مع رؤية المملكة 2030.

الكلمات المفتاحية: الذكاء الاصطناعي؛ الكتابة الأكاديمية؛ ChatGPT؛ Grammarly؛ المملكة العربية السعودية؛ مراجعة منهجية.