

The Role of Artificial Intelligence in Enhancing Scholarly Research – AI Tools Evaluation

Sharon Qi, San Jose State University

Xiaohong Iris Quan, San Jose State University

Taeho Park, San Jose State University

Bobbi Makani, San Jose State University

Abstract

This paper investigates the role of artificial intelligence (AI) in academic research, for example, ChatGPT, Research Rabbit, LitMaps, Scite AI, Elicit, and Copilot. Understanding their impact on research processes is crucial as AI technologies become increasingly sophisticated. This study identifies fifty generative AI software tools and their primary functions for scholars. It examines how these tools differ in functionality complexity, accuracy, reliability, and validity in academic contexts. Through empirical data collected from experiments conducted by the researcher team, the study assesses the effectiveness of fifty AI tools that may potentially assist academic research.

The findings of this research contribute to a deeper understanding of how AI tools can enhance scholarly productivity, streamline research processes, and potentially reshape the future of academic work. By offering practical insights and recommendations, this study aims to inform scholars, educators, and institutions about the opportunities and challenges associated with incorporating AI into academic research.

Keywords: Artificial intelligence (AI), AI tools evaluation, AI in scholarship, Research process efficiency

Introduction

Artificial Intelligence (AI) has started to revolutionize various fields in academia, where AI-powered tools are increasingly utilized to enhance research productivity and writing efficiency. According to the Economist report, 10% or more of abstracts for papers in scientific journals now appear to be written at least in part by large language models; In fields such as computer science, that figure rises to 20% (Economist, 2024). According to another new study by Harvard Business School, when AI is used by highly skilled workers, it can improve a worker's performance by as much as 40% compared with workers who don't use it (Somers, 2023). AI has recently been defined as "the use of computational machinery to emulate capabilities inherent in humans, such as doing physical or mechanical tasks, thinking, and feeling" (Huang and Rust, 2021, p.31). More and more AI tools have emerged for academic use. The application of AI tools in academia ranges from literature reviews, data analysis, to abstracts compiling and writing aids, such as grammar check and reference management. The promise of AI goes beyond simple automation. It also includes the potential to extract patterns and insights from large volumes of data that are sometimes invisible to human researchers and can even synthesize new research ideas, allowing researchers to focus more on critical thinking and sophisticated problem-solving.

The academic community has shown a keen interest in exploring the effectiveness of specific AI tools, especially since the release of ChatGPT in November 2022, with the primary focus predominantly being on individual tools like ChatGPT (Aydın, 2023; Dwivedi, Kshetri, Slade, Jeyaraj, and Kar, 2023; Hosseini, Rasmussen, and Resnik, 2023; Nguyen-Trung, Saeri, and Kaufman, 2023; Salvagno, Taccone, and Gerli, 2023). But there are a few exceptions with focus on other AI tools. For instance, Kurniati and Fithriani (2022) examined the use of Quillbot for enhancing English academic writing, and Marzuki, Widiati, Rusdin, Darwin, and Indrawati (2023) investigated a few AI writing tools on student writing quality from the perspective of English as a Foreign Language (EFL) teachers and demonstrated the benefits of integrating some AI writing tools for EFL students. However, there remains a significant gap in the literature concerning a comprehensive and systematic evaluation of the wide variety of AI tools available for academic work.

This evaluation study aims to address this gap by providing a comprehensive evaluation of multiple AI tools designed for academic use. By identifying and assessing the capabilities, benefits, and limitations of approximately fifty AI tools, this study seeks to offer valuable insights into how these tools can enhance the academic research and writing process. Such an evaluation is crucial for academics to make informed decisions about how to integrate AI technologies into their research work, ultimately advancing scholarly productivity and innovation.

Literature Review

The scope of the literature review in this section focuses on studies on AI software tools for scholars, ranging from AI tools that assist literature review, academic writing, and publication review, to other writing-related assistance such as reference management, as well as concerns about the reliability of AI tool, such as ChatGPT, and the related ethics.

AI assisting literature review

For scholars to conduct research, a literature review usually serves as one of the initial steps. AI's potential to expedite literature reviews has been explored in several studies, although the use of AI in this context is in an early stage of development.

Many studies have highlighted the potential positive role of AI in the research process, although concerns about reliability have been repeatedly raised. Wagner, Lukyanenko, and Paré (2021) provided a comprehensive overview of AI's capabilities in this area, which reviewed how the reported AI applications prior the release of ChatGPT were applied to a set of steps of the literature review process, for example, steps of problem formulation, literature review, searching, screening for inclusion, quality assessment (i.e., methodological flaws and source of bias), data extractions (i.e., getting relevant information), and data analysis & interpretation (i.e., either descriptive syntheses or inductive work such as theory generation), as well as proposing a research agenda for AI-based literature reviews (AILRs) at three different levels (i.e., supporting infrastructure, methods and tools, and research practice). They highlighted AI's efficiency in handling large volumes of documents and facilitating literature synthesis. Johnson, Bauer, and Niederman (2021) mentioned a tool called ORA, which is a dynamic network analysis tool for analyzing Scopus or other online bibliographic sources (e.g., Ebsco, Google Scholar, Orcid, ProQuest, Web of Science). Unlike two studies reviewed here that were published pre-Chat GPT age, Nguyen-Trung, Saeri, and Kaufman (2023) demonstrated how AI tools like ChatGPT and ChatPDF could enhance evidence reviews, despite limitations such as inconsistent results and certain errors.

Although these AI-powered tools demonstrated some comprehension of research concepts, it has been repeatedly reported that the AI tools sometimes misinterpreted material or generated misleading descriptions or summaries of those concepts. For example, in the test of using ChatGPT for literature review in the field of medicine, Haman and Školník (2023) found that in only 17 out of 50 instances, the articles could be located within the databases (such as Google Scholar, PubMed, Semantic Scholar), while 66% of the references produced by ChatGPT were non-existent papers.

At the methodological level, Guler, Kirshner, and Vidgen (2023) highlighted the effectiveness of combining machine learning and ChatGPT in literature reviews, noting improvements that have made in identifying research topics and generating content. While they used machine learning to identify research topics, ChatGPT was used to assist in labeling the topics, generating content, and improving the efficiency of academic writing.

AI assisting academic writing

AI has the potential not just to improve both the efficiency of literature reviews but also the overall productivity of academic writing. Significant exploration has been conducted into how certain AI tools can be effectively leveraged to enhance the academic writing process, as well as to understand their limitations.

Abdul, Mathew, Ahmad Saad, and Alqahtani, (2021) provided a comprehensive review of AI's role in scholarly writing. They identified tools across various categories, including literature

search and review, writing and editing, references and citations, review and workflow, plagiarism checking, and journal selection. However, since their paper was published in 2021, before the release of ChatGPT, like some other reviewed papers, the tools they identified do not encompass the recent advancements in AI within this field.

Khalifa and Albadawy (2024) conducted another comprehensive review study on AI-assisted academic writing. After an initial screening of 217 studies, they filtered their selection further down to 24 studies. Through analysis, they identified six core domains where AI supports academic writing, including 1) idea development and research design, 2) content development and structuring, 3) literature review and synthesis, 4) data management and analysis, 5) editing, review, and publishing support, and 6) communication, outreach, and ethical compliance. They summarized each of these 24 papers by their titles, main focus, key findings, what the AI application is about, limitations, and recommendations. After they mapped these 24 studies with the six domains, they found that domain 1) was the least researched, while domains 5) and 6) were mostly researched. In their conclusion, they addressed the importance of integrating AI tools into the academic research process, while they mentioned the ethical and transparent use of AI tools, as well as the need of training in using AI tools. They urged people to continue researching AI's impact on academic research since this is an ever-evolving process. They also made a few recommendations for AI technology in assisting research including “developing advanced AI tools for hypothesis formulation and predictive analysis”, and “establishing ethical frameworks for AI use” (p10).

Most studies on AI for scholars in the current literature have predominantly focused on ChatGPT, an AI tool known for its comprehensive features, particularly in academic writing. A notable study in this area is the opinion research conducted by Dwivedi, Kshetri, Slade, Jeyaraj, and Kar (2023). The study involved many authors. They extensively investigated the impact of ChatGPT through 43 contributors across fields including computer science, marketing, information systems, education, policy, hospitality and tourism, management, publishing, and nursing. The evolution of AI technologies was overviewed, and the positive impact of ChatGPT on various industries was acknowledged. The authors raised concerns such as “threats to privacy and security, and consequences of biases” (p3). They suggested further research in three areas along with many detailed research directions: 1) “knowledge, transparency, and ethics”; 2) “digital transformation of organizations and societies”; and 3) “teaching, learning, and scholarly research.” (p3)

Salvagno, Taccone, and Gerli (2023) discussed the utility of AI chatbots like ChatGPT in scientific writing, assisting researchers and scientists in organizing material, generating an initial draft, and/or in proofreading. They evaluated tools such as Elicit and compared the performance of AI chatbots with that of humans. They concluded that while AI tools were useful, they did not surpass humans in highly technical areas, particularly in selecting the most appropriate wording. They also emphasized that human oversight was crucial to ensure accuracy and prevent plagiarism. This aligned with Haman and Školník's (2023) findings on the limitations of AI tools in generating accurate academic references.

Ray (2024) highlighted the potential of ChatGPT in early career research scholarship, mentioning its role in refining hypotheses and conducting literature reviews. However, the discussion should have delved deeper into these aspects.

Marzuki, Widiati, Rusdin, Darwin, and Indrawati (2023) conducted a qualitative study examining the influence of AI writing tools on student writing. He interviewed four English as Foreign Language (EFL) teachers who had years of EFL teaching experience and a certain time of using AI tools as part of their teaching curriculum. Based on his research review on AI's impact on teaching writing, he designed an interview instrument and collected opinions and feedback from these participants. Each teacher used 3-4 different AI tools, including Quillbot, WordTune, ChatGPT, Essay Writer, PaperPal, and Jenni where QuillBot and WordTune were used by four different teachers. The overall impression was that these AI tools improved the content and organization of student writing, although the level of potential positiveness of AI tools to students' English writing skills was different. This qualitative study provided a rich insight into EFL teachers' teaching experiences using AI tools.

AI impact on the publication industry

In addition, AI-supported technologies are rapidly changing the publishing industry including reviewing and editing, which is another key domain integral to the research process. For instance, AI-driven software like ChatGPT, Grammarly, and Paperpal can correct grammatical errors and improve writing style. Tools like Zotero, Mendeley, and EndNote are indispensable for literature management. Turnitin and Copyscape stand out in the domain of plagiarism detection, employing extensive databases to verify the originality of academic works. The UK Publishers Association investigated the role of AI in the publishing industry, including AI investments and the obstacles faced by the sector. Their findings were published in a report titled *The Role of Artificial Intelligence in Publishing* (2020), which revealed that most publishing sectors believe AI "will be important over the next five years" (p3). The report also noted that "AI investment in the sector has just begun," with "larger publishers leading the drive" (p3).

AI-assisted writing education

Researchers have also explored the impact of AI on students and the educational process. A study of 343 communication instructors revealed a collective view that AI-assisted writing is widely adopted in the workplace and requires significant changes to instruction despite challenges such as less critical thinking and authenticity in writing (Cardon et al, 2023). For instance, Cribben and Zeinali (2023) reviewed the benefits and limitations of ChatGPT in business education and research, noting its utility in designing courses, creating content, and grading. Steele (2023) argued that AI tools like ChatGPT could empower students by enhancing their comprehension, research, and composition skills. Nevertheless, he also noted the AI-related threats to traditional education systems, such as measurement problems and skill devaluation. Mishra (2024) investigated the integration and experiences of academic professionals with AI tools in their pedagogical practices and illustrated a broad understanding and adoption of AI tools. The findings highlighted the need for adequate training and ethical guidelines for responsible AI use.

Kurniati and Fithriani (2022) studied how post-graduate students viewed Quillbot as a digital tool for English academic writing by employing a qualitative case study design involving 20 post-graduate students majoring in English education. The findings revealed that the postgraduate students in the study responded positively to using Quillbot to assist them in improving the quality of their writing.

Reliability and ethical concerns

Mandai, Tan, Padhi, and Pang (2024) highlighted the propensity of AI tools to generate content based on statistical probabilities rather than understanding, leading to errors and hallucinations. Dashti, Londono, Ghasemi, and Moghaddasi (2023) tested whether ChatGPT could find same articles in *The Journal of Prosthetic Dentistry (JPD)* as people searched directly in the journal and Google Scholar using a set of same keywords at different time. They found the results did not match the papers that ChatGPT had generated. Furthermore, all 75 articles provided by ChatGPT were not accurately located in the JPD or Google Scholar databases.

In an editorial for *the Journal of Accountability in Research*, researchers (Hosseini, Rasmussen, and Resnik, 2023) explored the performance of ChatGPT in writing tasks. Their findings revealed that the chatbot sometimes produced responses that were either entirely incorrect or not pertinent to the given topics. Consequently, they advocated for a rigorous review process, recommending that "any section of a manuscript created by an NLP system should be meticulously examined by a domain expert to ensure its accuracy, detect any biases, maintain relevance, and evaluate the reasoning presented (p1)."

Ciaccio (2023) addresses the necessity of recognizing AI's role in scientific writing. Basic AI assistance, such as spell-checking and grammar correction, typically does not require acknowledgment and can be effectively managed by AI tools. However, more extensive editorial interventions, including content editing—such as reorganizing paragraphs or sections, rewriting passages, and adding or deleting content—should be transparently disclosed by authors when submitting manuscripts for publication.

Aydın (2023) explored the use of Google Bard for generating literature reviews, comparing its performance to ChatGPT. They found both tools showed promise but exhibited higher plagiarism rates and occasional inaccuracies, underscoring the need for careful monitoring and ethical guidelines.

According to research by Casal and Kessler (2023) on *Applied Linguistics journals*, the journal reviewers were largely unsuccessful in identifying AI versus human writing, and many editors believed there are ethical uses of AI tools for facilitating research processes. Further research is needed to address the matter.

Research Questions

In summary, the literature as reviewed above presented both the potential benefits and the challenges of AI and AI-based tools in academia. These tools can potentially enhance research productivity and writing quality significantly, although concerns of inaccuracy and ethical warnings were consistently addressed. We found that there was very little systematic and in-depth research conducted regarding AI-assisted academic tools' quality evaluation, such as

accuracy, reliability, and scalability, as well as usability. This is a critical issue as we embrace AI in the academic research process, and it warrants careful consideration. To address this gap and help provide comprehensive guidelines in using AI-assisted academic tools, we conducted a study with the following primary goals and research questions. We present the results in the later part of this explorative paper.

The primary goal of this paper is to explore the emergent AI tools that can potentially assist scholars in doing their academic research more effectively and efficiently, using a thorough evaluation methodology that supports them in making a sound decision on what AI tool they should use for different purposes.

Research Question 1: What generative AI software tools are currently available for academic scholars, and how do they differ in their primary functions and complexity?

Research Question 2: How do these AI tools vary in terms of accuracy, reliability, validity, usability, and scalability in assisting scholarly work, as well as the associated costs?

Research Question 3: What are the key recommendations in helping researchers decide on which AI tool may fit what needs, and along what concerns regarding the adoption and integration of AI tools in academic research?

Methodologies

Participants

A student research team of eight, who majored in software engineering or computer science, six as undergraduates and two graduates, supervised by the first two authors, participated in data collection. The team was sponsored by school faculty research funding and school early career training funding. Weekly meetings were held for coaching, data collection review, data cleaning, and data organizing.

Data collection procedure

We processed the data collection by three steps below:

Step 1: We started a pilot testing with one AI tool for 3 weeks. Based on the pilot experiences, we figured out an effective way to collect data.

Step 2: Once we created a best practice guideline and defined the scope of exploration, we moved on to exploring a large set of AI-assisted research tools. We investigated these tools from the research-related functions' perspective, such as searching and finding references, processing literature review, judging reference's relevance and references/citations' mapping, helping paper writing idea creating and structuring, assisting wording selection, and proofreading.

Step 3: We further filtered and shortened the tool list, and re-evaluated them regarding accuracy, reliability, and validity. Then, we explored the complexity and usability of the tools. The ranking was based on a rigid rating criterion. See the details below. The data was processed using a cross-tester rating, and the average was calculated to mitigate the subjectivity.

The data collection process started in November 2023 and ended in June 2024.

Initial data sources

To systematically identify and select AI tools relevant for scholarly use, we employed a comprehensive approach that involved multiple diverse arrays of reputable data sources. The initial phase of our data collection involved sourcing potential AI tools from websites dedicated to technology and education, online rankings of AI tools, peer-reviewed academic papers, expert blogs, and credible news outlets. The selection criteria were rigorously defined to ensure that only legitimate and well-regarded tools were considered. Specifically, tools that had been consistently highlighted in the literature for their innovative capabilities, widespread usage, or endorsements by experts in the field were given precedence. We also cross-referenced these tools with online reviews and user feedback to gauge their effectiveness and reliability in academic settings. A list of over 80 AI tools were initially identified.

Data sources filtering

This initial list was subsequently refined through a careful selection process that aimed to ensure a balanced representation of AI tools catering to different aspects of scholarly work. The final selection comprised 50 AI tools, which were chosen based on their ability to perform key functions critical for academic productivity. These functions include searching for papers on specific topics, reading and analyzing academic papers, providing summaries, checking searched sources' relevance, reference mapping, citation assistance, content organizing, and assisting with academic writing. The selected tools span a range of capabilities from advanced search engines that utilize AI to pinpoint relevant academic papers, to sophisticated text analysis tools that offer in-depth comprehension aids and writing assistants that support the drafting and editing process. The diversity of the selected tools ensures that they collectively address the broad spectrum of tasks that scholars typically engage in during their research and writing processes.

Cross-tester rating

These filtered 50 AI tools were ranked based on specifically given criteria. See the next section for details. Each selected tool was evaluated by two testing team members (evaluation participants). Their rating average was calculated to mitigate subjectivity. See the final matrix in the research findings section later in this paper.

This evaluative process was designed to capture the user experience and practical utility of each tool from the perspective of end-users, namely, how each participant would most benefit from these technologies. By incorporating user feedback into our methodology, we aimed to provide a holistic assessment of the AI tools, balancing technical functionality with practical usability. The resulting rankings offered valuable insights into which tools were most effective in supporting academic scholarship and which features were mostly valued by users in an academic context. This methodological approach, grounded in both extensive literature review and empirical user feedback, ensured that the selected AI tools are not only technically proficient but also practically relevant to the needs of scholars. The careful selection process, combined with rigorous evaluation by the research team, provided a robust foundation for understanding the

current landscape of AI tools available for academic purposes and their potential impact on scholarly productivity.

Rating factors/ranking scale

We outlined a detailed protocol used for our research team to evaluate the filtered 50 AI tools across a range of critical metrics. The selected tools were rigorously tested based on eight key criteria: cost, accuracy, reliability, validity, function complexity, usability, efficiency, and scalability. The criteria guided a comprehensive assessment of each tool's performance and explored not only their effectiveness but also their adaptability to the diverse needs of academic work. See below for details.

Cost. The cost of using each AI tool was a consideration in our evaluation process. We examined both upfront costs and ongoing expenses, including subscription fees and any additional charges associated with premium features or updates. The affordability of each tool was assessed relative to its functionality, to determine whether the cost is justified by the tool's capabilities and benefits. We did not use a Likert scale for the cost factor but provided direct cost information based on three categories that we identified through the research. The three categories include free, freemium, or fee-based. The cost information would be able to accommodate the budgets of researchers, faculties, students, and institutions who need to plan their budgets in adopting them.

Accuracy, reliability, and validity. Accuracy is a crucial metric for AI tools, particularly in academic settings. We tested if the tool did exactly what it said (e.g., answer the questions correctly). We assessed the accuracy of each tool in performing specific tasks such as grammar checking, data extraction, and content generation. This involved testing the tools against a set of benchmark tasks where the expected outcomes were well-defined. For example, grammar-checking tools were evaluated based on their ability to identify and correct syntactical errors, while data extraction tools were tested on their precision in retrieving relevant information from complex datasets. The tools were scored based on their error rates and the relevance of the results they produced. We adopted the Likert scale of one through five for the ranking purpose. Reliability was measured by examining the consistency of each tool's performance over multiple use cases across different contexts. We tested if the tool performed the same function across three different instances (i.e., three different academic papers from the literature list provided). This involved repeated testing of the tools under varying conditions to ensure that they could deliver consistent results irrespective of the complexity of the task or the volume of data processed. A tool's reliability was determined by its consistency and ability to function without failure or significant performance degradation over time. Tools that demonstrated high stability and consistent output were ranked higher. Like reliability, reliability is essential to tools that scholars depend on for critical academic tasks.

The validity of each AI tool was assessed by determining the extent to which it measures what it claims to measure. We tested whether what was done was true, for example, whether the generated citation was truly existing. This involved comparing the tool's outputs with established standards or expert assessments to ensure that the tool accurately fulfilled its intended purpose. A tool's validity was considered high if its results closely matched the expected outcomes or expert judgments, indicating that the tool was effective in achieving its stated objectives.

Likert Scale	1) answer your questions exactly as expected;	2) complete the task with consistent quality across three or more different instances;	3) each verification proves the truth of the generated content (e.g., the citation is truly existing, or the data facts are true).
5	100%	100%	100%
4	80%	80%	80%
3	50%	50%	50%
2	30%	30%	30%
1	0%	0%	0%

Table 1: Likert Scale for Accuracy, Reliability, and Validity

Function Complexity. Function complexity refers to the range and sophistication of the functions provided by each AI tool. Tools were evaluated based on their ability to perform a diverse array of tasks, as well as the depth of customization and control they offered to users. We measured the complexity of a tool based on the basic scholarly functions, such as literature searching, relevance checking, paper reading (reviewing and analyzing), reference and citation, and paper writing.

Likert Scale	Function/Features	Criteria Description
5	<p><u>Functions:</u></p> <p>Searching</p> <p>Reading</p> <p>Writing</p> <p><u>Features:</u></p> <p>5-8 as listed on the criteria description</p>	<p>Functions: All 3 functions searching, reading, and writing are present</p> <p>Features: 5-8 or more features (e.g., finding paper, checking relevance, summary, detailed paper analysis, editing, chatting, citation, references, related paper finding, recommending the writing outline, etc.).</p> <p>Part of a suite of tools: The tool is linked to another tool, not isolated, as a part of a suite, e.g., ChatPDF vs ChatGPT. Each</p>

Likert Scale	Function/Features	Criteria Description
	<p><u>Part of a suite of tools?</u></p> <p>Yes</p>	<p>claimed feature is effective, and the database is also comprehensive.</p>
4	<p><u>Functions:</u></p> <p>1-2 functions of those listed above</p> <p><u>Features:</u></p> <p>1-4 as listed on the criteria description</p>	<p>Functions: 1-2 functions (only reading and writing, or just searching, with a limited database or a special database)</p> <p>Features: 1-4 features (editing, citation, etc.) provided very thoroughly.</p> <p>Part of a suite of tools: The tool is relatively isolated or stands alone, but it performs one function/feature effectively. The effectiveness exceeds the tool with the same function/feature in the more comprehensive tool suite that is scaled 5</p>
3	<p><u>Functions:</u></p> <p>1-2 functions of those listed above</p> <p><u>Features:</u></p> <p>1 as listed on the right, but unique</p>	<p>Functions: 1-2 functions (either reading or writing, or both, but no searching function at all)</p> <p>Features: Some limited feature(s) provided effectively, e.g., summarizing a paper only, or checking relevance only. But it is so unique that it cannot be substituted.</p>

Likert Scale	Function/Features	Criteria Description
2	<p><u>Functions:</u> 1-2 functions of those listed above</p> <p><u>Features:</u> 1 as listed on the right, not unique</p>	Functions: 1-2 functions and provides some limited features that can be easily substituted by other tools.
1	<p>Functions: 1 function of those listed above</p> <p>But it is very difficult to duplicate the result due to its technical roadblocks</p>	Functions: 1 function, and cannot perform reliably, e.g., it cannot be opened, or it requires a long time of waiting for authorization.

Table 2: Likert Scale for Function Complexity

Usability. Usability was assessed by evaluating the ease of use and the user interface (UI) design of each tool. This metric focused on how intuitive and accessible the tool was for researchers who had various levels of technical expertise background. We analyzed the overall user experience. Tools that provided a seamless and user-friendly experience, with clear navigation and minimal learning curve, were rated higher in usability.

Likert Scale	Time needed	Criteria Description
5	<1 min	Users can figure it out how to use the tool in a few seconds (<1 min) and find buttons for different features in 1-5 minutes.
4	10-15 min	The tool has a 10 to 15-minute learning curve
3	16 -30 min	The tool requires taking a tutorial training to learn it for 30 minutes or more
2	>30 min	The tool is very hard to use, and no user guides are provided at all
1	> 1 hour	The tool simply doesn't work, after an hour of engagement with it

Table 3: Likert Scale for Usability

Efficiency and scalability. Efficiency was measured by the time and resource consumption of each tool in performing tasks. We evaluated how quickly each tool could complete assigned tasks compared with the traditional manual way. Tools that performed tasks swiftly were ranked higher, as efficiency is critical in academic settings where time may be limited. Additionally, we considered how well the tools managed large datasets and complex calculations, with no compromising the accuracy and reliability.

Scalability refers to a tool's ability to handle increasing amounts of work or data without compromising performance. Tools that demonstrated the ability to scale effectively, handling higher volumes of data or more complex tasks without significant loss of performance, were rated higher. Scalability is particularly important for tools used in research environments where the scope of work may expand over time. A Likert scale was used and calculated with a percentage, which was the recorded time processed with the AI tool divided by the recorded time processed by transitional manual way to reach a rating from one to five. The unit of time recorded was in seconds.

Likert Scale	Criteria in %	Criteria Description
5	25-39% or below	A Likert scale was used and calculated with a percentage, which was the recorded time processed with the AI tool divided by the recorded time processed by transitional manual way to reach a rating from 1 to 5. The unit of time recorded was seconds. If it is 5, it means that they are very different, and the AI tool is dramatically faster and more efficient across multiple testing instances.
4	40-54%	It means that they are quite different, and the AI tool is more than twice more efficient than the traditional tool.
3	55-69%	It means that they are different, and the AI tool is faster, but not more than twice as efficient as the traditional tool.
2	70-84%	It means that they are somewhat different, and the AI tool is slightly better
1	85-100%	It means that they are somewhat similar.

Table 4: Likert Scale for Efficiency and Scalability

Findings

Figure 1 showed how the filtered 50 AI tools that we investigated were classified by these functions:

- Read (9 in total),
- Read & Write (8 in total)
- Search, Map, & Citation (10 in total)
- Search & Read (2 in total)
- Search, Read & Write (8 in total)
- Write (13 in total)

Based this classification, we built a Table 5 series, in which each category belonged to one function, for example, Table 5(a) for Read, Table 5(b) for Ready & Write, Table 5(c) for Search, Map (i.e., Mapping), & Citation, Table 5(d) for Search & Read, Table 5(e) for Search, Read & Write, and Table 5(f) for Write.

At the bottom of each subtable (i.e., Table 5 series, or Table 5(a), Table 5(b), Table 5 (c), Table 5 (d), Table 5 (e), and Table 5 (f), you will see sub-total evaluation scores by these indexes:

- Accuracy
- Reliability
- Validity
- Complexity
- Usability,
- Efficiency and Scalability
- Total score

From the individual tool's perspective, we aggregated evaluation scores in the very far right side column of each sub-table 5 series (i.e., Tables 5(a) - Table 5 (e)) into Table 5(f). Here, 32 out of 50 or 64% of the tools were scored at 24-29 (here 30 as the total scores). Among them, 14 of them, or 28% of the tools were scored at 27-29 (out of 30). Three of the tools that notably stood out and nearly achieved full scores were ChatPDF, Copilot, and BingAI, due to their higher scores in the factor of complexity. This data indicates that AI tools that have rich features and good complexity is more favored.

Table 5(f), as a summary table, indicated that 39 out of 50 or 78% were scored at four or above for **accuracy**, 41 out of 50 or 82% at four or above for **reliability**, 42 out of 50 or 84% at four or above for **validity**, 27/50 or 54% at four or above for **complexity**, 38 out of 50 or 76% at four or above for **usability**, and 32 out of 50 or 64% at four or above for **efficiency and scalability**. This data indicates that more AI tools performed stronger in reliability and validity, with the accuracy seemingly acceptable overall, but fewer AI tools were able to provide comprehensive features.

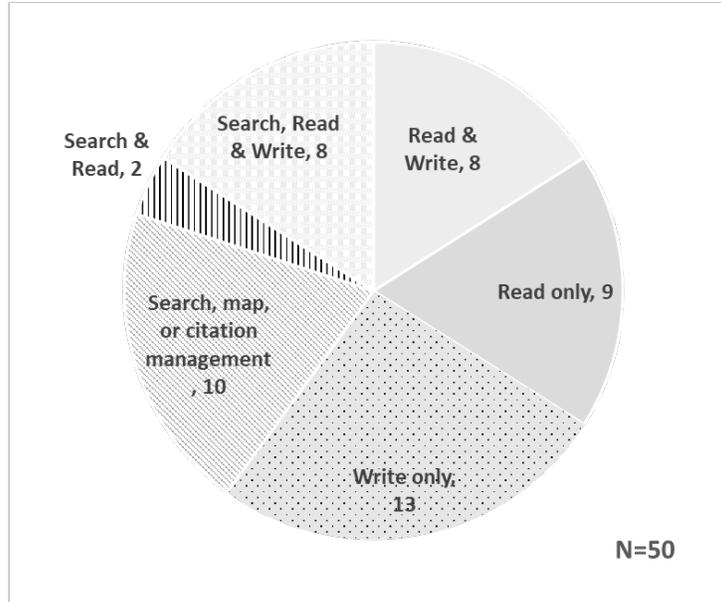


Figure 1: AI Tools for Scholars by Function Types

AI Tool Name (N=50)	Major Function (s)	Cost	Accuracy	Reliability	Validity	Complexity	Usability	Efficiency & Scalability	Total Scores by each tool
Mindgrasp AI	Read	Fee-based	3	3	3	3.5	4.5	4.5	21.5
Citation.ai	Read	Free	4	3.5	5	2.5	5	4.5	24.5
Summarize Paper	Read	Free	4.5	4.5	4.5	2.5	3.5	4.5	24
Resoomer	Read	Free	3	3	4	3	3.5	4	20.5
PDF AI	Read	Free	4.5	4.5	4.5	3.5	4.5	4.5	26
Myreader	Read	Freemium	3.5	4	4.5	3	4.5	4.5	24
Docalysis	Read	Freemium	4.5	4.5	4.5	4	4.5	4.5	26.5
SciSummary	Read	Freemium	3.5	3.5	3.5	3	4	4.5	22
Typeset	Read	Freemium	3.5	4.5	4.5	3.5	4.5	4.5	25

Table 5 (a): 50 AI Tools for Scholars by Function Types

AI Tool Name (N=50)	Major Function (s)	Cost	Accuracy	Reliability	Validity	Complexity	Usability	Efficiency & Scalability	Total Scores by each tool
Scholarcy	Read, Write	Fee-based	4	4	4	4	4	4	24
PDFgearCopilot	Read, Write	Free	5	5	5	4	5	3	27
Chatgpt	Read, Write	Freemium	4.5	5	3.5	4.5	5	5	27.5
ChatPDF	Read, Write	Freemium	5	5	5	4.5	5	5	29.5
Sider AI	Read, Write	Freemium	3.5	3.5	3	5	3.5	4.5	23

Scispace	Read, Write	Freemium	4	4	3	5	4	5	25
ScholarAI	Read, Write	Freemium	5	4.5	5	4	4	4	26.5
Bit.ai	Read, Write	Freemium	4	5	5	4	5	4	27

Table 5 (b): 50 AI Tools for Scholars by Function Types

Table 5 (c)									
AI Tool Name (N=50)	Major Function (s)	Cost	Accuracy	Reliability	Validity	Complexity	Usability	Efficiency & Scalability	Total Scores by each tool
EndNote	Search, Map, or Citation	Fee-based	5	5	5	2	3	3	23
zotero	Search, Map, or Citation	Free	4.5	5	4.5	3	2.5	1.5	21
Readcube	Search, Map, or Citation	Fee-based	5	5	4	4	4	1	23
Semantic Scholar	Search, Map, or Citation	Free	5	5	5	1	5	2	23
Mendeley	Search, Map, or Citation	Free	5	5	5	2	4.5	2.5	24
SourceData	Search, Map, or Citation	Free	1	1	1	1	2	1	7
Connected Papers	Search, Map, or Citation	Freemium	5	5	5	3.5	4	2	24.5
Research Rabbit	Search, Map, or Citation	Free	5	5	5	4	3.5	5	27.5
Citation Gecko	Search, Map, or Citation	Free	4.5	4.5	4	2.5	2.5	2	20
Litmaps	Search, Map, or Citation	Freemium	5	5	5	4	4	5	28

Table 5 (c): 50 AI Tools for Scholars by Function Types

Table 5 (d)									
AI Tool Name (N=50)	Major Function (s)	Cost	Accuracy	Reliability	Validity	Complexity	Usability	Efficiency & Scalability	Total Scores by each tool
Semantic Reader	Search, Read	Free	4.5	4.5	4	3	3.5	3	22.5
Scinapse	Search, Read	Freemium	3.5	4	4	2	5	3	21.5
Scite AI	Search, Read, Write	Fee-based	5	4.5	5	4	4	5	27.5
Elicit	Search, Read, Write	Fee-based	5	4.5	5	4.5	4	5	28
Meta AI	Search, Read, Write	Free	4	4.5	4.5	4.5	4.5	4.5	26.5
Consensus	Search, Read, Write	Freemium	5	5	5	4.5	4.5	4.5	28.5
Copilot Sidebar	Search, Read, Write	Freemium	5	5	5	5	5	5	30
Bing AI (Bing Copilot)	Search, Read, Write	Freemium	5	4.5	4.5	5	5	5	29
Gemini	Search, Read, Write	Freemium	4.5	4	4	5	4.5	4	26
Claude	Search, Read, Write	Freemium	4.5	4.5	4.5	4.5	4.5	3.5	26

Table 5 (d): 50 AI Tools for Scholars by Function Types

AI Tool Name (N=50)	Major Function (s)	Cost	Accuracy	Reliability	Validity	Complexity	Usability	Efficiency & Scalability	Total Scores by each tool
Word AI	Write	Fee-based	3.5	4.5	4.5	3	4.5	4.5	24.5
AI Writer	Write	Fee-based	4.5	3.5	4.5	4.5	5	5	27
ProDream Inc.	Write	Freemium	4.5	5	5	4	5	4.5	28
Lightkey	Write	Freemium	1	1	1	1	1	1	6
Grammarly	Write	Freemium	4.5	5	4	3.5	5	4.5	26.5
Scinote	Write	Freemium	3	3	3	4	2	3	18
Trinka	Write	Freemium	4.5	4	4	3.5	4.5	4	24.5
Crimson.ai	Write	Freemium	4	4	3	3	4	3	21
Junia	Write	Freemium	4.5	4.5	4	4	5	4.5	26.5
writesonic	Write	Freemium	4.5	4.5	4.5	4	5	4.5	27
Quillbot	Write	Freemium	5	5	5	4	5	5	29
jenni.ai	Write	Freemium	4	4.5	5	3	4.5	2.5	23.5
paperpal	Write	Freemium	5	5	5	4.5	3.5	2	25

Table 5 (e): 50 AI Tools for Scholars by Function Types

AI Tool Name (N=50)	Major Function (s)	Cost	Accuracy	Reliability	Validity	Complexity	Usability	Efficiency & Scalability	Total Scores by each tool
	All function types	Total Scores by Each Criteria	39	41	42	27	38	32	<i>32, 64%/50 tools (Scored 24-29)</i>
		Total % out of 50	78%	82%	84%	54%	76%	64%	<i>14, 28%/50 tools (Scored 27-29)</i>

Table 5 (f): 50 AI Tools for Scholars by Function Types

Looking into the price models, among the same filtered 50 tools that we evaluated (see Figure 2), 29 out of 50 or 58% were freemium (i.e., free at the beginning, following up with a fee charge), 13 out of 50 or 26% free, and 8 out of 50 or 16% fee-based, normally ranging from \$10 per month to \$40 per month. Such data indicates that the AI tools that have the writing feature are more likely to have a fee-charging. See Figure 2 below for more details.

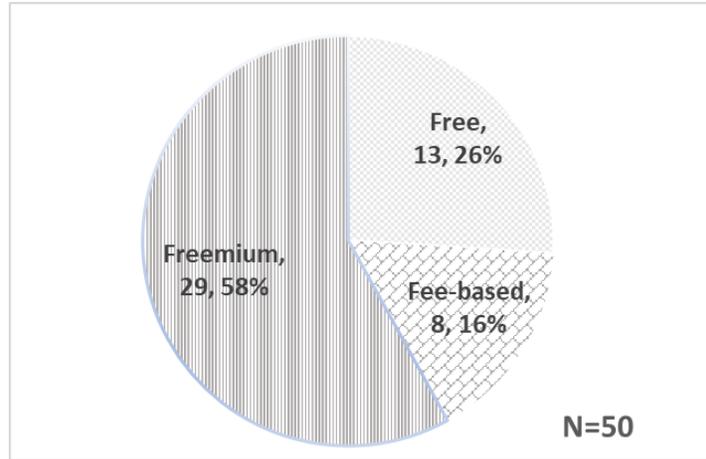


Figure 2: AI Tools for Scholars by Pricing Models

If we re-classified the same 50 AI tools by their price models (see Table 6), nine of the *read-only* type of AI tools had no significant differences in their accuracy, reliability, validity, complexity, usability, and efficiency/scalability. The eight *read-and-write* types of AI tools and those of the free price model appeared weaker in their accuracy, reliability, validity, complexity, usability, and efficiency/scalability than the other two models, particularly those with the freemium model. For ten of the *searches, map, or citation* type of AI tools, as well as two of *search and read* type of AI tools, they lose points for not having function complexity. For eight of the *searches, read and write* the type of AI tools, along with 13 of the *write-only* type of AI tools, they all scored relatively high, where the AI tools of the fee-based model stood out. This data indicates that the AI tools are still having challenges in improving their quality in reading and writing, while the AI tools may offer higher quality assistance if users pay for them.

AI Tools by Function Types and Fee Models	Accuracy	Reliability	Validity	Function Complexity	Usability	Efficiency & Scalability
Read	3.78	3.89	4.22	3.17	4.28	4.44
Fee-based (1)	3	3	3	3.5	4.5	4.5
Free (4)	4	3.88	4.5	2.88	4.13	4.38
Freemium (4)	3.75	4.13	4.25	3.38	4.38	4.5
Read, Write (8)	4.38	4.5	4.19	4.38	4.44	4.31
Fee-based (1)	4	4	4	4	4	4

Free (1)	5	5	5	4	5	3
Freemium (6)	4.33	4.5	4.08	4.5	4.42	4.58
Search, Map, or Citation (10)	4.5	4.55	4.35	2.7	3.5	2.5
Fee-based (2)	5	5	4.5	3	3.5	2
Free (6)	4.17	4.25	4.08	2.25	3.33	2.33
Freemium (2)	5	5	5	3.75	4	3.5
Search, Read (2)	4.00	4.25	4.00	2.50	4.25	3.00
Free (1)	4.5	4.5	4	3	3.5	3
Freemium (1)	3.5	4	4	2	5	3
Search, Read, Write (8)	4.75	4.56	4.69	4.63	4.5	4.56
Fee-based (2)	5	4.5	5	4.25	4	5
Free (1)	4	4.5	4.5	4.5	4.5	4.5
Freemium (5)	4.8	4.6	4.6	4.8	4.7	4.4
Write (13)	4.04	4.12	4.04	3.54	4.15	3.69
Fee-based (2)	4	4	4.5	3.75	4.75	4.75
Freemium (11)	4.05	4.14	3.95	3.5	4.05	3.5

Table 6: 50 AI Tools Evaluation Rating (scale 1-5) by Functions and Price Models

We also researched briefly the large language models used to support these AI tools. It appeared that most of the reviewed AI tools are supported by the GPT model (v3 free or v4 fee-based), which was the first AI LLM that went to the public, while Germin has its own LLM called

Gemini Pro 1.5, and Meta AI supported by its own LLM called Llama (v3.1). When people process text-based tasks, ChatGPT may be more preferred, while people process multimedia content or long sentence requests, Gemini may be more favored (Masaklkhhi, Ong, Waisberg, & Lee, 2024). Although Meta AI (Llama) came to the public late and only one year ago, when people process images or content related to multi-media, Meta AI (Llama) can help to provide up to 100 free images per day, in addition to its compatible functions in reading, writing and searching based on its huge amount of data sources. As Gemini integrated with its applications such as Google Chrome, Gmail, and Google Drive, Meta AI also integrated Meta AI with its social media applications such as Facebook and Instagram and with the latest information. Therefore, it is hard to comment on which LLM works better than the other (Timonera, 2024) since they all have different strengths. Besides, although GPT LLM stands alone, it provides APIs that many other applications can be integrated. On our evaluation list, other than Gemini and Meta AI, all the rest are integrated with and supported by GPT LLM.

Discussions, Conclusions, and Implications

As presented earlier, our findings addressed each of our research questions. Based on the evaluation of 50 AI tools for academic research, several recommendations and implications can be drawn. Firstly, our findings reveal a wide array of generative AI tools available to academic scholars, including the most current ones. These tools vary in their core functions, complexity, and ability to assist with tasks such as literature reviews, academic writing, and citation management. While this variety offers researchers numerous options to explore and select AI tools that best suit the different stages of academic writing, researchers should prioritize a needs-based approach to tool selection, aligning specific functionalities with research requirements. Objective evaluation criteria, encompassing accuracy, reliability, validity, complexity, usability, and efficiency/scalability, should guide this selection process. While complex, multi-functional tools, exemplified by ChatPDF, Copilot, and Bing AI proved advantageous for comprehensive research tasks, single- or dual-function tools such as Bit.ai, Research Rabbit, and Elicit demonstrated efficacy in addressing specific research needs.

Secondly, the evaluation data demonstrated that many AI tools performed well in terms of reliability and validity, with acceptable accuracy across various platforms. However, only a few AI tools offered comprehensive features. We identified three primary pricing models: freemium, fee-based, and free. Of these, the freemium model appeared to be the most widely adopted. In such cases, cost considerations warrant significant attention. The study revealed that the freemium model represents the most predominant pricing structure, but researchers should recognize that fee-based models frequently offer enhanced quality and more extensive feature sets.

Finally, ethical implications surrounding AI utilization in academic writing necessitate careful consideration. Transparency in AI deployment and adherence to evolving journal practices regarding using AI tools like ChatGPT, ChatPDF, Copilot Sidebar, Bing AI, and Gemini in publishing are paramount. These AI tools were noted for their rich functionality, particularly ChatPDF and Copilot, which excelled in performing internet searches and responding to prompts in an interactive, conversational manner, producing valid and accurate outputs. As highlighted earlier, three tools—ChatPDF, Copilot, and Bing AI—stood out, receiving near-

perfect evaluation scores largely due to their higher complexity. However, this does not imply that the other tools were of lower quality. A range of single- or dual-function tools also distinguished themselves with high scores in accuracy, reliability, or validity, and were noted for their ease of use. Examples include Bit.ai, Research Rabbit, LitMaps, Scite AI, Elicit, PDFgear, AI Writer, Pro Dream, and Quiltbot. These tools offer excellent performance based on specific user needs. For instance, Elicit, powered by multiple models, allows users to create custom prompts and presents results in a clear, tabular format. This functionality is especially valuable for querying multiple papers simultaneously, greatly facilitating the research process.

Our findings suggest that AI tools with more complex features are generally more preferred for assisting with academic writing, although some single- or dual-function tools also proved helpful for specific academic tasks. Cost is another factor to consider when selecting an AI tool, with the freemium pricing model proving to be more popular than fee-based or free models. When selecting AI tools, users should carefully consider their primary research needs and select tools according to their needs, using evaluation scores as we presented here or other published best practice as a guide.

As shown in our literature review earlier, only a limited number of studies have evaluated the effectiveness of AI tools in a comprehensive scope, and very little studies evaluated AI tools' quality from user-end perspective objectively. They either provided overview of other researchers' results in AI (e.g., Wagner, Lukyanenko, and Paré, 2021; Abdul, Mathew, Ahmad Saad, and Alqahtani, 2021; Khalifa and Albadawy, 2024) or focused on users' perspective about their subjective opinions (e.g., Marzuki, Widiati, Rusdin, Darwin, and Indrawati, 2023; Dwivedi, Kshetri, Slade, Jeyaraj, and Kar, 2023). Thus, this study has made a significant contribution to addressing this gap by providing not just comprehensive review but also an objective evaluation. Additionally, with the release of ChatGPT in November 2022, many studies published prior to that date have become outdated. This paper updates existing research by including newer AI tools not evaluated in those studies prior ChatGPT release, such as Wagner, Lukyanenko, and Paré (2021) study.

Furthermore, this paper enhances our understanding of how most current AI tools can improve the efficiency of the academic research process via a very structured evaluation. Particularly, based on what we found as presented earlier, we are very confident to claim that this study has enriched the studies about how newly published AI tools are changing the research academics landscape, how these various AI tools can potentially assist academic work more effectively and efficiently, and how we may evaluate their effectiveness carefully and constructively when we adopt these AI tools into our daily research work.

In addition, as explained earlier, each AI tool was rigorously tested using carefully designed metrics or guidelines, with results meticulously documented for further analysis. The evaluation methods included both quantitative measures (e.g., a 1-5 Likert scale for performance benchmarks and error rates) and qualitative feedback from users (e.g., ease of use and satisfaction levels). Also, the data was collected cross-tester for each evaluation task. This comprehensive approach provided a well-rounded assessment of each tool's strengths and weaknesses, offering a robust basis for comparison and evaluation. Such rigorous methodology supported the claim that the findings are valuable.

Recommendations

It is important to clarify that our assessment of reliability, validity, and accuracy focused primarily on the performance of specific tasks assigned to the AI tool—such as summarizing articles, providing in-depth analysis, or checking grammar—rather than verifying whether the papers identified by the AI tool actually existed, as highlighted and studied by Haman and Školník (2023). Their concerns about low accuracy in this regard merit further investigation, which could be the subject of a separate evaluation study. We hope this clarification prevents any misunderstanding regarding the scope of our accuracy-related evaluations.

Although we made efforts to mitigate bias, including averaging cross-tester scores, the subjective nature in deciding evaluation ratings could still pose a risk of bias. Therefore, we recommend future research to re-evaluate the AI tools as we assessed to either confirm or expand upon our findings. Future research should focus on validating the research findings through further investigation, refining evaluation methodologies, and critically examining the broader ethical ramifications of AI integration within academia. We hope that our evaluation methodologies will become more mature, after further verification and investigation, when it guides end-users and fellow researchers in selecting, assessing, and using appropriate AI tools to support their academic research.

Lastly, we want to emphasize the importance of addressing ethical concerns when using AI tools, although this is not the major focus of this study. We completely agreed that it is crucial to remain transparent about the extent of AI use when writing academic papers. Plagiarism through using AI assistance in writing has been a common concern in both academic research and educational practice. We encourage researchers to use AI-research tools mindfully and stay informed about the policies many academic journals have already implemented regarding AI use in scientific writing and publishing (Ciaccio, 2023). This is an important area that deserves further research.

Disclaimer

The authors of this paper declare that they have no commercial relationships or financial sponsorships with any of the AI tools discussed in this study. All evaluations and analyses presented are conducted independently, and the findings are based solely on the empirical data and research methodologies employed. The authors have no vested interest in promoting or endorsing any specific AI tool or company.

While writing this paper on AI tools for scholars, a few AI tools, including ChatGPT, were utilized to assist with the wording choices or grammatical checking occasionally. However, the primary arguments and findings are entirely grounded in the authors' research, conducted with the assistance of a group of student research team. All conclusions have been rigorously validated by the authors to ensure accuracy and integrity.

References

- At least 10% of research may already be co-authored by AI (2024, June 26). *The Economist*. Retrieved from: <https://www.economist.com/science-and-technology/2024/06/26/at-least-10-of-research-may-already-be-co-authored-by-ai>.
- Abdul Razack, H. I., Mathew, S. T., Ahmad Saad, F. F., & Alqahtani, S. A. (2021). Artificial intelligence-assisted tools for redefining the communication landscape of the scholarly world. *Science Editing*, 8(2), 134-144. <https://doi.org/10.6087/kcse.244>
- Aydın, Ö. (2023). Google Bard Generated Literature Review: Metaverse. *Journal of AI*, 7(1), 1-14. <https://doi.org/10.61969/jai.1311271>
- Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age. *Business and Professional Communication Quarterly*, 86(3), 257-295. <https://doi.org/10.1177/23294906231176517>
- Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, 2(3). <https://doi.org/10.1016/j.rmal.2023.100068>
- Ciaccio, E. J. (2023). Use of artificial intelligence in scientific paper writing. *Informatics in Medicine Unlocked*, 4(1), 101253. <https://doi.org/10.1016/j.imu.2023.101253>
- Cribben, Ivor and Zeinali, Yasser, The Benefits and Limitations of ChatGPT in Business Education and Research: A Focus on Management Science, Operations Management and Data Analytics (March 29, 2023). Available at SSRN: <https://ssrn.com/abstract=4404276> or <http://dx.doi.org/10.2139/ssrn.4404276>
- Dashti, M., Londono, J., Ghasemi, S., & Moghaddasi, N. (2023). How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing? *The Journal of Prosthetic Dentistry*. <https://doi.org/10.1016/j.prosdent.2023.05.023>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Guler, N., Kirshner, S., and Vidgen, R. A literature review of artificial intelligence research in business and management using machine learning and ChatGPT (August 14, 2023). UNSW Business School Research Paper Forthcoming, Available at SSRN: <https://ssrn.com/abstract=4540834> or <http://dx.doi.org/10.2139/ssrn.4540834>

- Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. *Accountability in Research - Ethics, Integrity and Policy* 1–9. <https://doi.org/10.1080/08989621.2023.2168535>
- Haman, M., & Školník, M. (2023). Using ChatGPT to conduct a literature review. *Accountability in Research*, 1–3. <https://doi.org/10.1080/08989621.2023.2185514>
- Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30-50. DOI:10.1007/s11747-020-00749-9
- Johnson, C. D., Bauer, B. C., & Niederman, F. (2021). The Automation of Management and Business Science. *Academy of Management Perspectives*, 35(2), 292–309. <https://doi.org/10.5465/amp.2017.0159>
- Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 5(1), 100145. <https://doi.org/10.1016/j.cmpbup.2024.100145>
- Kurniati, E. Y., & Fithriani, R. (2022). Post-graduate students' perceptions of Quillbot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics*, 7(3), 437-451. <https://doi.org/10.21462/jeltl.v7i3.852>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2). <https://doi.org/10.1080/2331186X.2023.2236469>
- Mandai, K., Tan, M. J. H., Padhi, S., & Pang, K. T. (2024). A cross-era discourse on ChatGPT's influence in higher education through the lens of John Dewey and Benjamin Bloom. *SSRN*. <https://ssrn.com/abstract=4755622> or <http://dx.doi.org/10.2139/ssrn.4755622>
- Masaklkhhi, M., Ong, J., Waisberg, E., & Lee, A.G. (2024). Google DeepMind's Gemini AI versus ChatGPT: a comparative analysis in ophthalmology. 38 (1412–1417), retrieved from <https://www.nature.com/articles/s41433-024-02958-w>
- Mishra, H. R. (2024). Exploring the impact of artificial intelligence tools in engineering pedagogy: A qualitative survey of academic experiences. *International Journal for Research in Applied Science and Engineering Technology*, 12(1), 60-66. <https://doi.org/10.22214/ijraset.2023.57864>
- Nguyen-Trung, K., Saeri, A. K., & Kaufman, S. (2023, April 20). Applying ChatGPT and AI-powered tools to accelerate evidence reviews. OSF Preprints, 20 April, 2023. <https://doi.org/10.31219/osf.io/pcrqf>

- Ray, P. P. (2024). Using ChatGPT for early career research scholarship. *The Clinical Teacher*, 21(3), 1–1. <https://doi.org/10.1111/tct.13694>
- Salvagno, M., Taccone, F.S. & Gerli, A.G. Can artificial intelligence help for scientific writing?. *Crit Care* 27, 75 (2023). <https://doi.org/10.1186/s13054-023-04380-2>
- Somers, M. (2023, October 19). How generative AI can boost highly skilled workers' productivity. *MIT Sloan Management Review*. Retrieved from <https://mitsloan.mit.edu/ideas-made-to-matter/how-generative-ai-can-boost-highly-skilled-workers-productivity>
- Steele, J. L. (2023). To GPT or not GPT? Empowering our students to learn with AI. *Computers and Education: Artificial Intelligence*, 5, 100160. <https://doi.org/10.1016/j.caeai.2023.100160>
- Timonera, K. (2024). Meta AI Chatbot Quick Facts, retrieved from eWeek at: <https://www.eweek.com/artificial-intelligence/meta-ai-chatbot-review/#:~:text=One%20feature%20that%20distinguishes%20the,15%20with%20Copilot's%20free%20plan>
- Wagner, G., Lukyanenko, R., & Paré, G. (2021). Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology*, 37(2), 209-226. <https://doi.org/10.1177/02683962211048201>

Acknowledgement

We extend our gratitude to the following student researchers for their invaluable assistance in collecting evaluation data for this research project:

Ajay Singh
Derrick Zihang Hong
Harika Kolipaka
Khaled Zaza

Mohammad Zaza
Sean Nian
Cody Ourique
Noel Hjalmarson

QRBD

QUARTERLY REVIEW OF BUSINESS DISCIPLINES

February 2025

Volume 11
Number 3/4



A JOURNAL OF INTERNATIONAL ACADEMY OF BUSINESS DISCIPLINES
SPONSORED BY UNIVERSITY OF NORTH FLORIDA
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)