dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240038

Quality and Effectiveness of AI Tools for Students and Researchers for Scientific Literature Review and Analysis

Martin DANLER[a,1], Werner O. HACKL[a] , Sabrina B. NEURURER [a,b] and Bernhard PFEIFER[a,b]

[a] *Division for Digital Medicine and Telehealth, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*
[b] *Department of Clinical Epidemiology, Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck, Austria*

Abstract. This study scrutinizes free AI tools tailored for supporting literature review and analysis in academic research, emphasizing their response to direct inquiries. Through a targeted keyword search, we cataloged relevant AI tools and evaluated their output variation and source validity. Our results reveal a spectrum of response qualities, with some tools integrating non-academic sources and others depending on outdated information. Notably, most tools showed a lack of transparency in source selection. Our study highlights two key limitations: the exclusion of commercial AI tools and the focus solely on tools that accept direct research queries. This raises questions about the potential capabilities of paid tools and the efficacy of combining various AI tools for enhanced research outcomes. Future research should explore the integration of diverse AI tools, assess the impact of commercial tools, and investigate the algorithms behind response variability. This study contributes to a better understanding of AI's role in academic research, emphasizing the importance of careful selection and critical evaluation of these tools in academic endeavors.

Keywords. Artificial Intelligence, Literature Analysis, Academic Writing

1. Introduction

The integration of Artificial Intelligence (AI) tools has significantly changed how students and researchers approach scientific literature in the ever-evolving landscape of academic research. This paper embarks on an exploratory journey to evaluate the quality and effectiveness of AI tools specifically designed to aid in scientific literature review and analysis. With an increasing reliance on AI to streamline research processes, it is imperative to understand the capabilities, limitations, and suitability of these tools for academic purposes [1].

Our research entailed a comprehensive investigation into various AI tools, focusing on those tailored for scientific research. The primary objective was to categorize these tools based on their functionality and effectiveness in assisting with different research related tasks. A critical aspect of our study involved creating a detailed table that

[1] Corresponding Author: Martin DANLER, UMIT TIROL, Hall in Tirol, Austria, E-Mail: martin.danler@umit-tirol.at

provides an overview of these tools, offering a clear and concise resource for students and researchers alike [1].

Our analysis was limited to freely available AI tools that allow users to pose direct research questions, in recognition of the accessibility challenges faced by many in the academic community. This constraint not only underscores the commitment to inclusivity but also reflects a realistic approach to resource availability for a significant portion of the academic community [2].

The following sections of this paper will discuss the methodology used to select and evaluate these tools, followed by the results and a detailed analysis of our findings. The objective of this work is to provide the academic community with valuable insights into the current state of AI tools in scientific research. This will enable informed decisions about their application in literature review and analysis. Through this endeavor, we hope to contribute to the broader discourse on the integration of AI in academic research, highlighting both its potential and its limitations.

2. Methods

The methodological approach adopted in this study was designed to systematically identify and evaluate AI tools suitable for scientific research and academic writing. This section outlines the steps taken to achieve a comprehensive overview of the available tools, culminating in a tabular presentation that categorises these tools based on their specific functionalities and applicability [3].

2.1. Internet Search and Keyword Selection

The primary method of data collection involved conducting extensive searches on Google, the world's most widely used search engine. To ensure a focused and relevant search, two sets of keywords were employed: "AI Tools for Scientific Research" and "AI Tools for Academic Writing." These keywords were chosen to capture a broad spectrum of AI tools that are specifically geared towards facilitating various aspects of academic research and writing.

2.2. Search Execution and Data Collection

Multiple searches were conducted using the selected keywords. Each search result was scrutinized for relevance, and only those websites, articles, and resources that directly addressed AI tools for scientific research and academic writing were considered for further analysis. This process involved evaluating the content of each source for its depth, accuracy, and applicability to the academic research context.

2.3. Tool Selection Criteria

The AI tools identified through the search were subjected to a set of selection criteria to ensure their relevance and utility for this study. The primary criteria included:

- **Accessibility:** The tool must be freely accessible, without requiring financial investment, thus ensuring inclusivity for all researchers and students, regardless of their institutional or financial resources.
- **Functionality for Direct Research Queries**: The tool must allow users to pose direct research questions or queries, a feature critical for academic research and literature analysis.

2.4. Tabular Representation and Categorization

A systematic table (Table 1) has been compiled of the AI tools found during the research. This tabular format was designed to provide a clear and concise overview of each tool, including its name, pricing model, primary functionalities that are suitable for specific research-related tasks, whether references are issued and whether a research question can be asked. The categorization aimed to assist researchers and students in easily identifying the tools most relevant to their specific needs.

Table 1. AI tools found during research.

Name	Pricing	Category	RQ	Ref
scite.at	Monthly fee	Search, Summarize, Analyze	No	Yes
Assistant by scite	Free (Beta)	Search, Summarize, Analyze	Yes	Yes
Iris.ai	Monthly fee	Visualize, Summarize, Manage, Analyze	No	Yes
Research rabbit	Free	Visualize, Summarize, Manage	No	Yes
Scispace	Free	Search, Summarize	Yes	Yes
ChatGPT	Free and monthly fee	Write, Summarize, Analyze	Yes	No
Consensus	Free	Search, Summarize	Yes	Yes
Elicit	Free and monthly fee	Search, Summarize	Yes	Yes
ChatPDF	Free	Summarize	No	Yes
Google Bard	Free (Beta)	Write, Summarize	Yes	Yes
Jenni AI	Free and monthly fee	Write, Summarize	No	Yes
Semantic scholar	Free	Search, Summarize	No	Yes
OpenRead	Free and monthly fee	Search, Summarize	No	Yes
Trinka	Free and monthly fee	Write	No	No
Microsoft Copilot	Free	Write, Summarize	Yes	Yes
Scholar GPT	Free	Search, Summarize	Yes	Yes

2.5. Analysis and Evaluation

Following the tools selection, each tool was evaluated for its effectiveness and quality in aiding academic research and writing. To scientifically evaluate these tools, we posed the following research question multiple times to each tool to test for consistency in the responses:

RQ: "How do Medical Data Warehouses influence the efficiency and quality of healthcare delivery?"

To ensure a thorough analysis of the AI tool outputs, our approach has expanded to include not only evaluating the outputs but also conducting a detailed assessment of their sources and volume.

Table 2 presents a systematic list of all AI tools investigated, along with the word count of their outputs and the number of sources they reference. Additionally, it indicates whether the cited sources are scientific in nature.

This enables a more nuanced analysis, offering a dual perspective on the content volume and source credibility — both critical factors in the evaluation of AI-assisted

literature review and analysis. Through this refined lens, we strive to present a clear and objective portrayal of each AI tool's capacity to support the expansive terrain of academic research.

The following section will explore the implications of these findings, explaining how the quantity of content and the quality of sources intersect to influence the usefulness of AI tools in academic contexts. This discussion aims to clarify the potential applications and benefits these technologies offer to the scholarly community.

Table 2. Outcome of each AI tool quantified.

Name	Word Count	References	Cited Sources
Assistant by scite	358	9	Yes
Scispace	153	5	Yes
ChatGPT	349	0	No
Consensus	328	5	Yes
Elicit	98	4	Yes
Google Bard	469	3	No
Microsoft Copilot	126	3	No
Scholar GPT	339	6	Yes

3. Results

The investigation into the effectiveness and quality of various AI tools for scientific research and academic writing yielded significant insights, primarily reflected in the diversity of responses to research queries. In our exploration of AI tools for scientific research, we uncovered a tapestry of variability in the responses to identical research questions. The outputs differed notably in scope and depth, a testament to the array of methodologies and algorithms at play within each tool. These differences pave the way for a multitude of interpretations and information presentations.

In delving into the provenance of information that the AI tools utilized, our analysis paid special attention to their source selection. Remarkably, the investigation revealed a nearly complete lack of shared sources among the tools, with one notable exception. The tools Elicit and Consensus stood out, as they both referenced the same source within their outputs. This was an isolated occurrence amidst a landscape where each AI tool otherwise appeared to access its own distinct repository of data. This singular overlap between Elicit and Consensus points to possible similarities in their design or source retrieval algorithms. Despite this instance of convergence, the general trend showed a wide-ranging array of sources across the tools, indicating a rich diversity of data inputs that contribute to the unique outputs each AI tool provides. This diversity highlights the breadth of information available to these tools and suggests a capacity for offering varied perspectives and insights, enriching the overall landscape of AI-assisted research.

When delving into the types of sources used, it became apparent that some tools, like Google Bard and Microsoft Copilot, were not limited to scholarly materials but also pulled from general websites and other non-academic domains. The integration of such materials raises important questions about the academic soundness and dependability of the output from these AI instruments.

In evaluating the AI tools, we repeatedly submitted the same research question to each system to assess the consistency of their responses. It was found that although the substance of the responses exhibited minor variations with each iteration, the word count of the provided answers remained strikingly similar across multiple queries. This suggests that while the AI's rephrasing capabilities introduce some degree of variability,

the underlying information processed by the tool is drawn from a stable set of sources. Furthermore, the constancy of the cited sources for each tool indicates a fixed reference database from which the AI retrieves information. This pattern of consistent word count and source usage provides an intriguing insight into the operational consistency of these AI tools, despite the superficial variation in their outputs.

Another intriguing finding was the temporal limitation in the data sourcing, particularly with tools like ChatGPT, which relied on information only up to April 2023. This raises concerns about the ability of such tools to provide the most up-to-date research findings or to reflect the latest advancements in a field.

Despite these variations, most AI tools were found to primarily utilize scientific sources. However, there was an observable opacity in how they selected the scientific papers they did use. This opacity could potentially introduce biases and calls into question the selection criteria for the information processed by these tools.

Among the findings, it was noteworthy that several AI tools provided responses of remarkable quality to the posed research questions, demonstrating a sophisticated understanding and synthesis of the available scientific literature. These high-quality responses showcase the potential of AI to contribute valuable insights and facilitate the initial stages of research.

However, it is crucial to underscore that despite the high quality of some responses, relying solely on AI-generated outputs for crafting a scientific paper is not advisable. The process of producing scholarly work involves critical analysis, interpretation, and a deep understanding of the subject matter, which extends beyond the capabilities of current AI tools. While these tools can serve as an effective starting point or aid in the research process, they cannot substitute the intellectual rigor and analytical depth required for academic writing. The variability in the quality and relevance of AIgenerated content further emphasizes the need for thorough review and supplementation with human expertise to meet the scholarly standards of scientific research.[4] In summary, while AI tools offer promising avenues for supporting literature review and analysis, our results indicate significant variability in their outputs and use of sources.

This variability highlights the need for careful consideration and scrutiny when using these tools for academic research purposes.

4. Discussion

This research has provided valuable insights into the capabilities and limitations of freely available AI tools in the context of scientific research and academic writing. However, it is important to acknowledge certain limitations of our study, which in turn suggest avenues for future research.

Our study had a significant limitation as we deliberately excluded commercial AI tools. This was due to our focus on freely accessible options. It raises the question of whether paid tools may offer better performance or more advanced features. To gain a more comprehensive understanding of the AI landscape in academic research, future studies should include a diverse range of commercial tools.

Our methodology was restricted to tools that allow for the direct input of research questions. We intentionally did not include commercial AI tools, focusing solely on freely accessible options. Future research should explore combinations of AI tools, such as pairing writing-focused tools with those specialized in analysis and research, to

uncover synergies and enhanced capabilities that may not be apparent when tools are used in isolation

Variability in the outputs of AI tools was another observation from our testing, with differences evident even when responding to identical queries. Future studies should delve into the reasons behind these discrepancies and their implications for reliability and usability in academic contexts.

Further complexity was observed in the responses to the same question posed by different users; despite largely consistent content, there was a variation in textual presentation and structure. Further research is warranted to investigate the algorithms and processing mechanisms that contribute to these variations, providing insights into how AI tools personalize responses and the extent to which this impacts the objectivity and consistency of the provided information.

Among the analyzed outputs, we found no critical perspectives toward medical data warehouses or their application in healthcare provision. All examined texts emphasized the benefits, such as improved healthcare efficiency, support for decision-making through data analysis, and advancement of research into new treatments. These positive outcomes are mirrored by contributions to cost reduction and improved operational efficiency, personalized patient care, and support for public health decisions and predictive analytics. While the potential and realized benefits of medical data warehouses are well-documented, the absence of explicit discussion around challenges, such as data privacy, data integration complexity, data quality limitations, or the need for careful implementation, points to a gap in the literature. Such critical perspectives are crucial for a balanced understanding of technology application in the real world and would encourage dialogue on necessary improvements, safety measures, and ethical considerations regarding the use of data warehouses in healthcare provision.

The potential of open-access AI tools in academic research has been clearly demonstrated in this study. As AI continues to advance, deepening our understanding of its applications and implications in academic research becomes increasingly important. Future research pathways are plentiful and should aim to broaden our knowledge of AI tools, enhance their functionality in academic research, and address the ethical and practical nuances of their use in real-world scenarios.

References

[1] A. Gabriel, Artificial intelligence in scholarly communications: An elsevier case study, *Information Services and Use* **39**(4) (2020), 319–333

[2] R. Buchkremer et al., The Application of Artificial Intelligence Technologies as a Substitute for Reading and to Support and Enhance the Authoring of Scientific Review Articles, *IEEE Access*, **7**. Institute of Electrical and Electronics Engineers Inc., 65263–65276, (2019),.

[3] J. I. Martínez-López et al., Which are the tools available for scholars? A review of assisting software for authors during peer reviewing process, *Publications*, **7**(3). MDPI AG, (2019),.

[4] M. Hosseini et al., Using AI to write scholarly publications, *Accountability in Research* (2023), 1–9