



AI literature review systems: an analysis of performance, affordances, and outputs for a complex topic in the social sciences

Heather Moulaison-Sandy, Wilson Castaño-Muñoz, Laura Ridenour, and Denice Adkins

DOI: <https://doi.org/10.47989/ir30iConf46906>

Abstract

Introduction. Artificial intelligence has the potential to revolutionize the way that scholars produce work, including through the creation of literature reviews using free or low-cost systems. Despite the importance of the literature review, AI-powered tools to create literature reviews are understudied at present.

Method. To address this gap, four specialized tools were assessed. The complex area of study chosen was e-reading in English and Spanish, a field where differences are evident in the two languages. Systems were prompted to write a literature review of 500-1000 words, and the resulting outputs were analysed based on the requirements for reviews of the literature.

Results. Although none of the systems allowed for all the identified criteria to be addressed, Scite allowed many criteria to be met and was able to produce text in Spanish as well as in English. The remaining three systems investigated, Jenni.ai, OpenRead, and Wisio, performed less favourably.

Conclusions. AI-powered systems show incredible promise, with one unstudied area being the creation of scholarly reviews of the literature. Although the kinds of research questions that might be examined in social sciences like LIS will necessarily be complex, until now little has been done to evaluate these systems.

Introduction

Artificial intelligence (AI) systems show promise in supporting scholars in the discovery and analysis of articles that are part of the scholarly record. A suite of AI-powered tools is emerging to support the creation of reviews of the literature. Specifically, these systems advertise the ability to analyse articles that will create new meaning, identify lacuna in the scholarly record, and support the understanding of the state of the field and, by extension, further steps that should be taken. By including scholarly sources, these systems allow researchers to explore the literature in their field, often for free or very low cost.

The performance, outputs, and affordances of these systems has only been evaluated anecdotally. Reddit threads have probed this question, in some cases with limited success (e.g., ijmarkandel, 2023). In other words, although the accuracy of aspects of certain systems (e.g., references in ChatGPT (Mugaanyi et al., 2024)) have been evaluated, very little is known about free and low-cost AI systems that support the creation of complex reviews of the scholarly literature, and no formal studies have been conducted.

In the social sciences, complexities readily emerge when studying the scholarly record relating to the position of the authors. Different terms can be used to describe a single phenomenon and different methodologies and literatures can be sources; language traditions also have different approaches to scholarship. Anecdotally, electronic reading (also known as e-reading, digital reading, or reading on screen) is one topic that is studied differently in anglophone and Hispanophone traditions. It represents a complex, social sciences-based problem that has not adequately been addressed in the literature.

This paper focuses on the novel topic of the performance of AI systems and their functionalities relative to traditional work in creating reviews of the literature in the social sciences, especially when working in novel, complex spaces. The complex topic explored is e-reading as studied in the scholarly literature in English and Spanish.

Area for application: e-reading

Some specific concepts in the field of digital reading have been covered using different terms. One such concept is the process by which a new medium reshapes an established one, simultaneously referencing and reinterpreting its conventions and traditions, termed remediation in English (Bolter & Grusin, 1996). Works in Spanish also address this concept in relation to the book and its evolution to the digital book, focusing on the materiality, format and features of the new medium. However, studies in Spanish do not use the term *remediación*, nor do they cite the English-language studies that coined the term (e.g., Canclini et al., 2015; Cavallo & Chartier, 1999; Cordon, 2016; Morales & Espinoza, 2003).

Given that with e-reading, researchers from different language families address similar concepts using different terminology and by citing different sources, understudied complexities emerge. To address this, we investigate the ability for free and low-cost AI-powered systems to create complex reviews of the literature. To our knowledge, although the topic of AI-powered tools and systems have been analysed for their usability in academia, no scholarly papers investigate tools for reviews of the literature by analysing their performance and affordances. The current project aims to address this gap in the literature.

Research question

To investigate this topic, the following research question is explored:

RQ1: To what extent are AI systems able to carry out sophisticated reviews of the literature for use in the social sciences?

Review of the literature

Literature reviews in the social sciences

In scholarship, reviews of the literature ground a project in the extant literature, situating it and allowing it to contribute to the field (Machi, L. A. & McEvoy, 2016). High quality literature reviews organize and synthesise completed work (Zhao et al., 2024). They help to identify a ‘gap’ and ideally are shaped to provide an argument as to what scholars in a given field should be able to do in the future (Slater, 2018). Practically, the sources will be carefully chosen, and elements including source, author, and date will play a role. The perspective adopted will support the research questions and will limit the scope of the work meaningfully, presenting an argument (Machi & McEvoy, 2016); in making choices to include sources, the extent of the analysis of the results and the scope of the discussion are likewise limited. Developing the ability to craft such an argument is a skill that takes years for a human researcher or scholar to develop (Joos et al., 2024; Slater, 2018).

AI-powered systems to support scholarship

AI-powered tools have emerged that support scholarship. This paper focuses on a set of AI-powered tools that allows for the automated creation of literature reviews or other kinds of analyses of scholarly articles. Because literature reviews are related to the discovery of articles and the connectedness of articles, these tools are related to and often overlap with other sets of AI-powered tools.

For example, to carry out a literature review, scholars must first identify relevant scholarly articles. One example of a searchable corpus is SemanticScholar (<https://www.semanticscholar.org>). The AI-powered search engine Consensus (<https://consensus.app/>) makes use of this corpus (Consensus AI, 2022). Scopus AI (<https://www.elsevier.com/products/scopus/scopus-ai>) has a similar approach. Other AI-powered search engines exist (e.g., Supersymmetry (<https://www.supersymmetry.ai/>)) that also support enhanced search and discovery. Lateral (<https://www.lateral.io/>) and Scispace (<https://typeset.io/>) support discovery and compiling notes.

Researchers can benefit from other support as well, and many systems have multiple ways of supporting authors. Another class of AI-powered tools is systems designed to help with the analysis and summarization of articles. For example, ChatPDF (<https://www.chatpdf.com/>), Explainpaper (<https://www.explainpaper.com/>), Humanata (<https://www.humata.ai/>), ResearchMate Pro (<https://researchmate.pro/>), Scholarcy (<https://www.scholarcy.com/>), Sharly (<https://sharly.ai/>), StudyFlo (<https://www.studyflo.com/>), and Unriddle (<https://www.unriddle.ai/>) support quick overviews of articles, in some cases if the publisher is willing to expose the source to analysis. Other AI-powered systems enable the discovery or visualization of connections between articles, including Connected Papers (<https://www.connectedpapers.com/>), Inciteful (<https://inciteful.xyz/>), Litmaps (<https://www.litmaps.com/>), and ResearchRabbit (<https://www.researchrabbit.ai/>).

Overview: AI-Powered systems supporting scholars and literature-based discovery

Artificial Intelligence Systems used for developing literature reviews are largely some forms of Large Language Model (Shopovski, 2024). One of the basic tenets of scholarly communication is the need for scholars to position their work in the field of endeavor. Goals of using LLMs is to streamline the process of surveying the literature (Joos et al., 2024), identify documents for inclusion in a review (Antu et al., 2023), evaluate and categorize relevant documents (Joos et al., 2024), interpret documents in the context of research, code relevant documents (Joos et al., 2024), and aid in the human-driven synthesis of relevant materials.

Free generative AI systems like ChatGPT (<https://openai.com/chatgpt/>) and Gemini (<https://gemini.google.com/>) are well-known for being good sources of inspiration, but their reliability, especially for scholarship where precision is essential, has been repeatedly questioned. For example, the phenomenon of hallucinations extends to sources, with ChatGPT (GPT-3.5) generating either incorrect DOIs or providing the wrong DOI for a source in 61.8% of natural sciences references and 89.4% of humanities ones (Mugaanyi et al., 2024). Hallucinations in AI systems can be divided into one of two types, 'intrinsic hallucinations' and 'extrinsic hallucinations' (Mittal et al., 2024). In the first type, intrinsic hallucinations, the output that is generated contradicts the source. In the second type, extrinsic hallucinations, the output is not able to be verified against the source, and is unverifiable.

As mentioned, a new suite of AI-powered tools has emerged, designed to support research. These tools have been organized into systems that help researchers find sources, make sense of primary source texts, and correct their own writing (e.g., Pinzolit, 2023). In addition, literature-based discovery (LBD) is a domain showing much promise. In this model, systems use AI to analyse the literature, specifically seeking lacuna and new approaches. LBD is generating interest in areas such as medicine (e.g., Cheerkoot-Jalim & Khedo, 2021).

Using AI-powered systems to create literature reviews for articles is understudied. Only two studies begin to address the problem. Choe and colleagues (2024) designed a study to assess novice researcher needs in creating reviews of the literature and then created LLM-based scaffolding for the work. Zhao and colleagues (2024) generated literature reviews using three AI-powered document summary systems and one literature review system; they applied Pattern Analysis and Machine Intelligence (PAMI) approaches, comparing AI-generated reviews to human-authored ones using new bibliometric indicators. The results identified limitations in the AI-powered outputs. To our knowledge, no systematic assessment of freely available and specialized AI-powered literature review tools or of their functionalities and their usefulness to academics has been carried out.

Method

Focusing on free and low-cost AI-powered systems that are designed and advertised specifically to support scholars in the creation of literature reviews and that are available to individual researchers (i.e., not requiring enterprise-level commitment by an institution such as with Laser (<https://www.laser.ai/>)), the research team identified 4 systems for analysis:

- Jenni.ai (<https://app.jenni.ai/>)
- Scite (<https://scite.ai/>)
- OpenRead (<https://www.openread.academy/>)
- Wisio (<https://wisio.app/>)

Each system was expertly learned by one team member who watched tutorials and practiced extensively, piloting each system. Based on requirements for literature reviews in LIS, a social science (e.g., Machi & McEvoy, 2016; Slater, 2018), and the affordances of AI-powered systems, deductive coding was used for this project.

Paid versions of each system were used to explore maximum benefits, and only the system's sources were used (no sources were uploaded). The goal was a usable 500 to 1000-word literature review on e-reading research in English and Spanish that 1) focused on both languages, 2) cited current and peer-reviewed sources, and 3) summarized methodologies; the research team's system's expert led a live demonstration for the research team beginning with the prompt: 'write a one-page literature review (500-1000 words) about e-reading in Spanish and English.'

Team leads then modified based on system affordances and results. During the demonstration, which was recorded, members of the research team collaboratively coded each system using an

inductive system based on the requirement of literature reviews in the social sciences, discussing each code until agreement was reached.

Results

Each AI-powered system designed for authoring reviews of the literature was investigated for its ability to produce text for a complex question. Although these systems functioned very differently, the results of their use and observations regarding their feedback, functionality, and responses are presented below.

Table 1 provides an overview of elements relating to authoring a literature review. Specifically, this project sought to produce more than 500 words on the topic, using recent peer-reviewed sources in both English and Spanish, and having the methodologies summarized in the analysis. Scite was the most performant of the group, creating a robust review of the literature using recent sources where methodology was described. None of the systems, however, allowed users to cite only peer-reviewed sources. Jenni.ai and Wisio were the least performant, with Jenni.ai unable to address the majority of the criteria, and only able to isolate methodology in a limited way. Wisio was also unable to address the majority of the criteria, although it was able to cite more recent sources when prompted, but not specifically by date or date range.

System name	>500-word review produced	Specify sources as peer reviewed	Specify sources by language	Specify sources by date	Isolate methodology
Jenni.ai	no	no	no	no	somewhat
Scite	yes	no	no	yes	yes
Open-Read	yes	no	yes	no	somewhat
Wisio	no	no	no	somewhat	no

Table 1. Overview of free and low-cost AI-powered systems for literature reviews.

Table 2 provides an overview of the sources cited and the formatting of the references produced in the test review. As the sources are an essential aspect of the literature review, this table focuses on foundational aspects of how they are provided as well as how the references are given and their accuracy. Again, Scite was the superior product, successfully addressing all the criteria listed. Each product linked to sources in some way or another and none produced hallucinations; Scite provided APA-style formatting and references, and the output text was downloadable as a file that included the formatted references. Jenni.ai likewise met these three criteria, but did not always provide references for sources mentioned in the text. OpenRead and Wisio provided links to sources, but did not provide a standard reference list and did not provide an easy way to download or save outputs. Secondary sources were at times provided in Jenni.ai, OpenRead, and Wisio in lieu of primary sources. The number of references provided ranged from 3 to 10.

System name	Link to sources	Primary sources only	All sources are referenced	Num. references provided	Hallucinations in references	Downloadable output with references	Able to output using APA
Jenni.ai	some	no	no	3	no	yes	yes
Scite	yes	yes	yes	10	no	yes	yes
Open-Read	yes	no	yes	3	no	no	no
Wisio	yes	no	no	10	no	no	no

Table 2. Sources, references, and output produced.

Additionally, scholars could upload sources themselves for inclusion to all four systems, although this project only used sources provided by the systems.

Observations about using the systems

The same prompt was used to begin each interaction with a given. However, the way that each system permitted interaction and prompt refinements differed. Jenni.ai and OpenRead were primarily prompt-based, functioning much like generative AI systems such as ChatGPT and Gemini. After the initial prompt, Jenni.ai allowed sentence-by-sentence approval of content as the primary way to add to the generated content, or for the selection of text around which additional content could be developed. Scite was the most like a traditional database, distilling prompts down to queries and then displaying the queries. Giving OpenRead an identity as a helpful assistant who speaks both languages produced a more scholarly review of literature, though the team observed that the content was not topically aligned with the prompt. Wisio seemed designed to assist in the writing process rather than replace it, with tools to check grammar and to align user input content with scientific writing styles, which Wisio calls ‘*scientifying*’ the writing.

Ability to refine the outputs

Each system allowed for different refinements to the text, via changes to the prompts (e.g., Scite), isolation of specific text to make edits (e.g., Jenni.ai). Features relating to language were difficult to isolate, in all systems producing either no or unpredictable results. For example, Jenni.ai and Wisio were unable to produce a critical mass of references in English, and efforts to request references in Spanish were likewise unsuccessful. Scite was able to change the output language to Spanish, which was unexpected since the prompt was provided in English. In OpenRead, clicking the follow-up option and using the same prompt produced a four-paragraph Spanish-language summary of e-reading research with several APA-style in-text citations. Further prompting was needed before the summary was translated to English and links to references were provided; however, those references were not consistent with the in-text citations.

Scite was the only system that allowed for a date range to be provided for references; however, Jenni.ai, OpenRead, and Scite all provided additional metadata and links to sources. Wisio seemed to use only PubMed and CrossRef to provide citations, but the AI Advisor feature did suggest other databases to search and keywords to use. The overall quality of references varied greatly. For example, OpenRead and Wisio were unable to format references in APA, and OpenRead did not allow for the text of the review to be downloaded with the references appended; instead, each reference needed to be downloaded individually.

Discussion

The results of this investigation provide insight into AI-powered systems for the creation of literature reviews. Based on the performance and functionalities as well as the usefulness of the outputs, we find that the systems performed unevenly.

Overall, based on the requirements of reviews of the literature in the social sciences (Machi & McEvoy, 2016; Slater, 2018), we find that Scite performed admirably by producing text that was both scholarly and well-referenced in alignment with expectations for reviews of the literature (Slater, date). Overall, the Scite output was not that far removed from the kind of output one might expect to find in a published article. Jenni.ai, OpenRead, and Wisio produced sub-par results, even after extensive massaging of the prompts and work within the system. Jenni.ai was unable to produce more than 500 words or include more than 3 references, though it mentioned additional sources in the body of the text. OpenRead produced a paragraph of text with several cited sources, but on review, included journal submission instructions in the content of the text. OpenRead also produced extensive references and scrolling produced more; however, the first 3 references were used in the text and the others were suggestions for related content. Wisio produced partial paragraphs, often ending with unfinished sentences. References provided were not specific to the

prompt, though Wisio's AI Advisor feature did describe and cite one study that met requirements for timeliness and that seemed to be on-topic.

All of the systems allowed for sources to be loaded and used. Uploading articles under copyright for use by these systems is possibly questionable in terms of intellectual property rights. Accordingly, this project sought to understand the affordances of the AI-powered systems without such inputs; however, for users who wanted to focus on specific sources, the ability to upload those sources might be a useful approach.

Limitations and future research

This exploratory study does not seek to answer questions about e-reading in English and Spanish empirically, nor to weigh in on the ethics of the design, functionality, or use of such systems in academic writing, but rather to compare the functionalities of free and low-cost AI-powered systems for creating literature reviews. The research questions investigated the nature of using the systems and the mechanics of the output, rather than the more subjective veracity of the contents. Further research will investigate the accuracy of the text produced, the relevance of the sources cited, and the ability of the AI to identify the most important and pressing concerns relating to the topic. As AI advances, further research should also continue to monitor the strengths and weaknesses of these and other systems.

Conclusion

AI-powered systems show incredible promise, with one unstudied area being the creation of scholarly reviews of the literature. Although the kinds of research questions that might be examined in social sciences like LIS will necessarily be complex, until now little has been done to evaluate these systems for their potential use in scholarship and teaching. This paper is a first attempt to systematically analyze the functionalities of four AI-powered systems for reviews of the literature that are available for use by individuals.

About the authors

Heather Moulaison-Sandy is Associate Professor in the iSchool, University of Missouri, USA. She received her Ph.D. from Rutgers University, and her research interests focus on the organization of information, scholarly communication, and technology. She can be contacted at moulaisonhe@missouri.edu

Wilson Castaño-Muñoz is Associate Professor at the School of Library and Information Science, University Antioquia, Colombia. He is a PhD student at the University of Missouri. His research interests are in digital reading and information conveyed through aural technologies. He can be contacted at wilson.castano@mail.missouri.edu

Laura Ridenour is Assistant Professor in the iSchool, University of Missouri, USA. She received her Ph.D. from the University of Wisconsin-Milwaukee, and her research interests include data science, artificial intelligence, information organization, and open data governance and access. They can be contacted at lerbhn@missouri.edu

Denice Adkins is Professor in the School of Information Science & Learning Technologies, University of Missouri. She received their Ph.D. from the University of Arizona and her research interests are in services to library users. She can be contacted at adkinsde@missouri.edu

References

- Antu, S. A., Chen, H., & Richards, C. K. (2023). Using LLM (Large Language Model) to improve efficiency in literature review for undergraduate research. *LLM@ AIED*, pp. 8-16.
- Bolter, J. D., & Grusin, R. A. (1996). Remediation. *Configurations*, 4(3), 311-358.
- Canclini, N., Gerber Bicecci, V., & López Ojeda, A. (2015). *Hacia una antropología de los lectores*. D - Ediciones Culturales Paidós. México.
- Cavallo G., & Chartier R. (Eds.). (1999). *A history of reading in the west*. Amherst: University of Massachusetts Press.
- Cheerkoot-Jalim, S., & Khedo, K. K. (2021). Literature-based discovery approaches for evidence-based healthcare: A systematic review. *Health and Technology*, 11(6), 1205-1217. <https://doi.org/10.1007/s12553-021-00605-y>
- Choe, K., Park, S., Jung, S., Kim, H., Yang, J. W., Hong, H., & Seo, J. (2024). Supporting novice researchers to write literature review using language models. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1-9. <https://doi.org/10.1145/3613905.3650787>
- Consensus AI. (2022, September 14). Consensus and Semantic Scholar Partner for Approachable Information. <https://consensus.app/home/blog/science-answer-app-consensus-and-article-aggregator-and-discovery-tool-semantic-scholar-announce-their-partnership/>
- Cordón, J. A. (2016). La lectura en el entorno digital: nuevas materialidades y prácticas discursivas. *Revista chilena de literatura*, (94), 15-38. <https://www.scielo.cl/pdf/rchilite/n94/art02.pdf>
- ijmarkandel. (2023). AI tools for literature review. r/PhD. Reddit. https://www.reddit.com/r/PhD/comments/17nqf59/ai_tools_for_literature_review/?rdt=52038
- Joos, L., Keim, D. A., & Fischer, M. T. (2024). Cutting through the clutter: The potential of LLMs for efficient filtration in systematic literature reviews. *arXiv Preprint arXiv:2407.10652*. <https://arxiv.org/pdf/2407.10652>
- Machi, L. A. & McEvoy, B. T. (2016). *The literature review: Six steps to success*. 3rd ed. Thousand Oaks, California: Corwin.
- Mittal, A., Murthy, R., Kumar, V., & Bhat, R. (2024, January). Towards understanding and mitigating the hallucinations in NLP and Speech. In: *Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD)*, pp. 489-492.
- Morales, O., & Espinoza, N. (2003). Lectura y escritura: coexistencia entre lo impreso y lo electrónico. *Educere*, 7(22), 213-222.
- Mugaanyi, J., Cai, L., Cheng, S., Lu, C., & Huang, J. (2024). Evaluation of large language model performance and reliability for citations and references in scholarly writing: Cross-disciplinary study. *Journal of Medical Internet Research*, 26, e52935. <https://doi.org/10.2196/52935>
- Pinzolit, R. F. J. (2023). AI in academia: An overview of selected tools and their areas of application. *MAP Education and Humanities*, 4(1), 37-50. <https://doi.org/10.53880/2744-2373.2023.4.37>
- Shopovski, J. (2024). Generative artificial intelligence, AI for scientific writing: A literature review. Preprints, 2024060011. <https://doi.org/10.20944/preprints202406.0011.v1>

Slater, T. F. (2018). Improving your argument by identifying a literature gap. *Journal of Astronomy and Earth Sciences Education*, 5(1), i-ii.

Zhao, P., Zhang, X., Cheng, M. M., Yang, J., & Li, X. (2024). A literature review of literature reviews in pattern analysis and machine intelligence. *arXiv preprint arXiv:2402.12928*.

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).