



JOURNAL OF INNOVATIVE RESEARCH (JIR)

ISSN: 2837-6706 (Online)

VOLUME 2 ISSUE 3 (2024)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

Integration of Artificial Intelligence (AI) into the Data Extraction Phase of a Scoping Review

Paige Maylott¹, Shaminder Dhillon¹, Dina Brooks¹, Sarah Wojkowski^{1*}

Article Information

Received: October 20, 2024

Accepted: November 29, 2024

Published: March 25, 2025

Keywords

*Artificial Intelligence, Data
Extraction, Large Language
Models, Methodology, Scoping
Review*

ABSTRACT

This paper describes how artificial intelligence (AI) was used to assist with the data extraction phase of a scoping review, specifically comparing different AI models and the accuracy of AI-assisted data extraction compared to human extraction. Scoping reviews map existing literature on a topic and are useful for complex or under-reviewed subjects. Integrating AI, particularly large language models, can enhance processing speed and data analysis. Three models, ChatGPT 3.5 and -4 (both developed by OpenAI) and Copilot (by Microsoft), were compared to identify the best model for AI-assisted data extraction. Adobe Acrobat Pro's Optical Character Recognition (OCR) feature and 'ChatGPT Splitter' were used to manage image-based content and large sections of data. A custom script was iteratively generated and implemented with the source material. AI-assisted extraction results were compared to text extracted by an independent reviewer. ChatGPT-4 was utilized to enhance efficiency and accuracy of data extraction from 234 sources. While human extraction was more specific with verbatim information, AI was faster and sometimes provided more nuanced understanding, averaging 20 minutes per source compared to one hour for human extraction. ChatGPT-4's superior text processing capabilities made it the optimal choice. While AI advancements have streamlined data extraction, human oversight remains crucial to ensure accuracy and address biases. This methodology is especially beneficial for smaller research teams and emphasizes the importance of structured prompts and rigorous review. Careful planning and oversight can mitigate risks, ultimately improving the quality and efficiency of the review process.

INTRODUCTION

Scoping reviews are a form of knowledge synthesis which aims to map existing literature on a particular topic or research question, identify key concepts, theories, sources, and gaps in research. Scoping reviews cover broader topics and provide a preliminary assessment of the potential size and scope of available research (Peters *et al.*, 2020). This method is particularly useful when the topic is complex or has not been comprehensively reviewed (Arksey & O'Malley, 2005). Conducting a scoping review involves several best practices to ensure rigor and reliability. These include clearly defining the research question, developing a detailed protocol, conducting a comprehensive literature search, selecting relevant studies, charting the data, and synthesizing and reporting the results (Levac *et al.*, 2010). High-level revisions to date on this methodology have focused on standardizing these steps and enhancing transparency, particularly through frameworks such as the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) (Peters *et al.*, 2021). The PRISMA-ScR provides a 22-item checklist that guides researchers through the process, ensuring all necessary components are included and reported clearly (Tricco *et al.*, 2018).

Artificial intelligence (AI) refers to the simulation of human intelligence processes by machines, particularly computer systems (Russell *et al.*, 2022). AI involves several key capabilities, including natural language processing, knowledge representation, automated reasoning, and

machine learning to adapt to new circumstances and to detect and extrapolate patterns (Vinuesa *et al.*, 2020). The concept of AI encompasses various approaches to emulate human thinking and behavior, such as the Turing test approach to assess acting humanly, the cognitive modeling approach to understand and replicate human thought processes, and the rational agent approach to create systems that act to achieve the best outcome given the current circumstances (OpenAI *et al.*, 2023; de la Torre-López *et al.*, 2023). As AI systems progress towards increasingly capable intelligence, the alignment of AI's objectives with human values becomes crucial to ensure they act in ways that are provably beneficial to humans. AI applications range widely, impacting fields from natural language processing and robotics to cognitive science and beyond (Russell *et al.*, 2022). Large Language Models (LLMs), a subset of AI, are trained on vast amounts of text data and can understand and generate human-like text or audio output based on the input they receive (OpenAI *et al.*, 2023). Examples of LLMs include OpenAI's GPT-3.5 and GPT-4, as well as Microsoft's Copilot.

In research, AI has been reported to assist with data analysis, abstract screening, literature reviews, optical character recognition for multi-language support, and generating hypotheses (de la Torre-López *et al.*, 2023). The potential benefits of AI and LLMs in research are substantial. They can process and analyze large datasets much faster than humans, identify patterns and insights which may be overlooked, and provide a

¹ School of Rehabilitation Science, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada

* Corresponding author's e-mail: wojkows@mcmaster.ca

consistent and unbiased approach to data extraction and synthesis (Frueh, 2023; Qlik, 2024). These capabilities are particularly valuable in scoping reviews, where the volume of literature can be overwhelming, and timely extraction of relevant data is crucial (Bolaños *et al.*, 2024). This paper describes how the authors integrated AI into the data extraction phase of a scoping review. Specifically, the authors employed AI-assisted extraction techniques as a substitute for one research assistant involved in the data extraction phase of a scoping review. All previous screening steps (i.e., title and abstract; full text) had been completed with two independent reviewers, with a third available to assist with decision making if consensus could not be achieved.

The integration of AI into the data extraction phase of scoping reviews represents a novel approach aimed at improving efficiency and accuracy of this stage. This integration was necessitated by the need to extract data from more than 200 sources while upholding quality control methods and streamlining the extraction of pertinent information from research papers. Traditionally, this task requires at least two reviewers working collaboratively to ensure the reliability of the extracted data. However, given the volume of sources and the time-sensitive nature of the review process, the incorporation of AI, along with Human-in-the-Loop (HITL) methodology, offered a promising solution to expedite this phase while maintaining quality (Duke, 2023; Alshami *et al.*, 2023).

The scoping review for which AI-assisted extraction was employed aimed to identify strategies in the literature intended to improve the accessibility of health professional programs for students with disabilities. The study is part of a broader program of research about inclusion and retention of students with disabilities in health professional programs. The study protocol has been published by JBI Synthesis (Dhillon *et al.*, 2024) and has been registered on OSF.

MATERIALS AND METHODS

Participant Information

The Accessibility Projects Support Coordinator (PM) developed and executed an AI-assisted extraction method. Before initiating this method, PM engaged in discussions with the rest of the research team (SD, DB, SW) to review the required data to be extracted and typical processes for data extraction.

Step 1: AI Models Comparison

Three AI models were evaluated for their effectiveness in data extraction—ChatGPT 3.5 and ChatGPT-4, both developed by OpenAI, along with Copilot by Microsoft. During this evaluation, various performance metrics were considered, including each model's capacity to process large amounts of inputted data (buffer size), how many inputted instructions it could store at one time (generation capacity), and quality of output (OpenAI *et al.*, 2023). Two sources included in the scoping review were used to

evaluate the models following the procedure outlined in the below steps; one source was less than 10 pages while the other was more than 20 pages in length.

The assessment revealed ChatGPT-4 outperformed its counterparts by providing a larger buffer size and a more generous daily generation limit, thereby enhancing the efficiency of the data extraction process. Conversely, Copilot had constraints in both buffer size and daily generation capabilities, which diminished its efficiency despite producing outputs of satisfactory quality. Copilot's constraints significantly limited its ability to handle long documents within a reasonable timeframe. ChatGPT 3.5, although available at no cost to users, was less effective due to its tendency to produce lazy responses, otherwise known as behavior drift or responses to user prompts that have degraded in quality over the lifespan of the AI (Chen *et al.*, 2023). Therefore, for the purposes of our research, ChatGPT-4 was identified as the optimal tool, as it struck a balance between attention to detail and operational efficiency, primarily due to its robust text processing capabilities, substantial input buffer, and overall efficiency. This model minimized the necessity for fragmented data inputs—a critical advantage, especially when processing extensive documents.

Step 2: Preparing the Sources

Of the 234 sources included in the scoping review, we extracted data using AI assistance from 157 sources. Preparatory measures were undertaken to ensure the material's readiness for analytical processing. This preliminary phase involved ensuring that all sources were in PDF or HTML format, followed by a meticulous manual review to identify sources which may require optical character recognition (OCR) remediation, such as image-based PDFs or image-heavy sources that prevent copying the text for AI processing. This step was followed by the organization of accumulated texts into alphabetically arranged file folders. These actions were critical for navigating the inherent challenges presented by the source materials and the constraints of current AI technologies.

Step 3: Review of Source Material

The source material for the project mainly consisted of PDF documents. These PDFs often contained image-based content, which made it difficult for the AI software to access and analyze the text. To address this issue, Adobe Acrobat Pro's Optical Character Recognition (OCR) feature was employed. This tool converted image-based text into a format that was both searchable and accessible, thus preparing the documents for the subsequent stages of data extraction.

Step 4: Text Splitting

Given the character limit constraints inherent to AI models, it became necessary to divide the text into smaller, more manageable chunks. For this purpose, a free custom tool, named the 'ChatGPT Splitter,' was utilized. This

tool facilitated the segmentation of large volumes of text, ensuring the coherence of the material was maintained and that the material remained compatible with the AI models for efficient processing.

Step 5: Script Development

To direct the AI model effectively in extracting the relevant information from the segmented text, a standardized script was developed by one of the authors (PM). This script included high-level instructions, illustrative examples, and a structured list of specific questions that aligned with the prompts in the extraction table formulated by the research team, tailored to meet the study objectives. Through iterative refinement, the script was optimized to enhance the accuracy of the extracted data and to minimize potential errors.

Ethical Considerations

To uphold ethical research practices, all extracted data were documented and attributed to the AI model used (i.e., ChatGPT-4) in the scoping review extraction table, in addition to and independent of the extraction completed by the research assistant. Having two sources for the extraction enhanced rigor, as per usual, and ensured verification and evaluation of the AI-assisted output. Another member of the research team (SD) compared the AI-extracted data alongside data extracted by the research assistant for accuracy and consistency. If discrepancies were identified, SD reviewed the original source and made a final decision on the data extracted. Human oversight, including the incorporation of Human-in-the-Loop (HITL) methodology, was integral for critical evaluation of the AI-generated outputs, leading to valid and reliable interpretations as research findings (Duke, 2023; Alshami *et al.*, 2023).

Another ethical consideration is that most LLMs are trained on information that is inputted from users. To protect the research process, including outcomes, the data collection feature on ChatGPT-4 was turned off prior to inputting source material.

RESULTS AND DISCUSSIONS

The purpose of utilizing AI-assisted extraction in our study was to increase the efficiency of data extraction. Thus, when we compared the outputs between AI and human extraction, we considered both the efficiency and accuracy of the work. AI-processed data are generated as text and presented to the user formatted into easily referenced, bolded headings. The data were copied and pasted to corresponding columns in our research extraction table.

Efficiency

To begin the AI-assisted extraction process, there were several steps, as described above, not required for human data extraction. Specifically, source material was reviewed and prepared using Adobe Acrobat Pro's Optical Character Recognition (OCR) feature. An hour was needed for this work as only a handful of sources

required this preparation. Then, GPT splitter was used to break the text into manageable chunks to upload into ChatGPT-4. This was a laborious task which involved copying and pasting the entire text into the online application, then copying and pasting each chunk back into ChatGPT-4. This process took approximately five hours. Finally, a script was developed to provide instructions to ChatGPT-4 for the data required from each source. This was an iterative process that required one hour to write the initial script, and then approximately two additional hours of refinement to produce aesthetically pleasing results which displayed consistent output.

However, once the preparation of source material and AI was complete, the extraction process itself was faster (i.e., AI finding the relevant information in each source and then PM inputting it into the extraction table). Typically, this extraction process took 20 minutes per source when done by AI, but one hour per source when done by a human. Overall, approximately 52 hours were spent on AI-assisted data extraction of 157 articles compared to 157 hours spent through traditional human extraction.

Accuracy

When reviewing the data extracted from each source by AI-assisted means versus a human, the outputs were very similar. Although the number of discrepancies was not tracked, occasionally, a difference between human and AI approaches to extraction was evident. Table 1 provides examples of the differences in data extraction between human and AI. The research assistant (human) provided data that was mostly verbatim from sources once they recognized the information as a strategy to improve the accessibility of health professional programs. In Example 1 of Table 1, the authors (Volino *et al.*, 2021) organized their discussion of the challenges and considerations for preparing pharmacy students with disabilities for experiential learning by using strategies as headings. The research assistant recognized these strategies and listed them as the output of extraction. The AI-generated data was more generic. When the same article was uploaded to ChatGPT-4, AI recognized the same strategies, but rather than repeating them, ChatGPT-4 provided a general statement about the contents of the source, which was less informative. In this example, the research assistant provided more specific information to address the research question.

However, there were instances in which the AI-assisted data was more accurate than the research assistant. In Example 2, the strategy for improving accessibility displayed more nuance by describing problem-based learning approaches combined with structured debates and flexible assessments (Foster, 2008). In comparison, the research assistant picked out the sentence that included the word "strategy" but did not recognize the different parts to this strategy, and the team agreed. In instances where data was more nuanced, ChatGPT-4 was better able to identify and describe information pertinent to the research question.

Discussion

The integration of AI technology in research methodologies has advanced significantly since the data extraction phase of this study. Contemporary LLMs now offer enhanced capabilities for processing entire documents and images without the necessity for text splitting, thus reducing the need for extensive pre-processing (Van Noorden, 2023). These advancements render aspects of our earlier methodology, such as the use of OCR and text splitting tools, largely obsolete. Nevertheless, our approach to structured prompt creation and the rigorous review of AI-extracted data remains pertinent. These steps ensure that AI outputs are reliable and aligned with the research objectives.

Despite these technological advancements, the requirement for human oversight persists. AI models, while powerful, may still overlook context-specific nuances or replicate bias inherent in AI training data (Nazer *et al.*, 2023). Therefore, incorporating a Human-in-the-Loop (HITL) approach ensures that AI outputs are vetted and validated. This combination of AI efficiency and human judgment optimizes the data extraction process, thereby enhancing overall quality of scoping reviews (Van Noorden, 2023).

The differing approaches taken for data extraction by AI versus the research assistant may also vary in studies where AI is not used (i.e., between reviewers). However, accuracy is not necessarily compromised. Traditionally, scoping reviews require two reviewers to extract data and a process to ensure any discrepancies between reviewers is

addressed (Tricco *et al.*, 2018). The AI-assisted approach may be particularly beneficial to smaller research teams with limited personnel available to complete the data extraction process efficiently.

The integration of AI in research methodologies continues to hold significant potential, particularly as AI technology evolves. Future research should explore the balance between AI automation and human oversight, focusing on optimizing workflows and addressing ethical considerations such as data privacy and bias mitigation. This ongoing exploration will be crucial in maximizing the benefits of AI while ensuring the integrity and accuracy of research findings.

Potential Limitations of AI in Research

Despite the potential for significant advantages, there are also limitations to the use of AI in research. One primary concern is the accuracy and reliability of AI-generated outputs. AI models are trained on existing data and may reproduce biases present in that data (Bender *et al.*, 2021). Additionally, while AI can handle large volumes of data, it may struggle with nuanced or context-specific information which requires human judgment (Bender *et al.*, 2021). Another limitation is the potential for errors in extracting ambiguous text or drawing context from images (Wang *et al.*, 2024). To mitigate these risks, human oversight is essential to verify and validate AI-generated outputs. Furthermore, ethical considerations such as data privacy and the transparency of AI processes must be addressed (Wang *et al.*, 2024).

Table 1: Examples of Differences between Data Extraction

Source	Strategy from Source	Extracted by AI	Extracted by Human
Example 1: Volino <i>et al.</i> (2021)	“The authors present a series of challenges and considerations for educators preparing pharmacy students with disabilities for experiential learning, with the following headings: Establishing supportive culture for accommodations; identifying key stakeholders early in the process; Aligning learning outcomes and performance expectations across learning environments; Monitoring and adapting individual accommodation plans; Ensuring the bridge between learning environments remains functional and intact.”	“The authors suggest methods for bridging skills-based accommodation needs to experiential settings, while managing accommodations in an evolving practice landscape.”	“1. Establishing a supportive culture for accommodation; 2. Identifying key stakeholders early in the process; 3. Aligning learning outcomes and performance expectations across learning environments; 4. Monitoring and adapting individual accommodation plans; 5. Ensuring the bridge between learning environment remains functional and intact.”

Example 2: Foster (2008)	“A teaching and learning strategy that incorporated flexibility in accommodating to some extent, the preferred learning styles of the student group...Problem based learning was initially introduced to the student group, followed by structured debate. The use of debate as a learning tool was audited, since existing assessment criteria could be modified for its use.” (Page 33)	“The paper discusses implementing varied teaching and assessment methods to accommodate students with dyslexia, including problem-based learning, structured debate, and altering assessment criteria to be more inclusive and sensitive to different learning needs.”	“A teaching and learning strategy that incorporated flexibility in accommodating to some extent, the preferred learning styles of the student group.”
--------------------------------	---	--	---

CONCLUSION

The integration of AI into scoping review methodology has the potential to allow research teams to process large amounts of information in a timely manner. This increased processing speed could help ensure scoping review results are published promptly or when the review still accurately reflects the available literature. By leveraging the capabilities of advanced AI models like ChatGPT-4, research teams may streamline the extraction process, reduce the time required, and maintain high standards of accuracy and reliability. While the use of AI presents certain limitations and ethical considerations, careful planning and rigorous oversight can mitigate these risks, ultimately enhancing the quality and efficiency of scoping reviews.

REFERENCES

- Alshami, A., Elsayed, M., Ali, E., Eltoukhy, A. E., & Zayed, T. (2023). Harnessing the power of CHATGPT for automating systematic review process: *Methodology, case study, limitations, and Future Directions. Systems*, 11(7), 351. <https://doi.org/10.3390/systems11070351>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bolaños, F., Salatino, A., Osborne, F., & Motta, E. (2024). Artificial Intelligence for Literature Reviews: Opportunities and challenges. *Artificial Intelligence Review*, 57(10). <https://doi.org/10.1007/s10462-024-10902-3>
- Chen, L., Zaharia, M., & Zou, J. (2023, October 31). How is CHATGPT's behavior changing over time? *arXiv.org*. <https://arxiv.org/abs/2307.09009>
- de la Torre-López, J., Ramírez, A., & Romero, J. R. (2023). Artificial Intelligence to automate the systematic review of scientific literature. *Computing*, 105(10), 2171–2194. <https://doi.org/10.1007/s00607-023-01181-x>
- Dhillon, S., Roque, M. I., Brooks, D., & Wojkowski, S. (2024). Strategies to increase accessibility for students with disabilities in health professional programs: a scoping review protocol. *JBI Evidence Synthesis*, 22(12), 2625–2635. <https://doi.org/10.11124/jbies-23-00484>
- Duke, T. (2023). Human-in-the-loop. *Building Responsible AI Algorithms*, 95–103. https://doi.org/10.1007/978-1-4842-9306-5_6
- Foster, I. (2008). Enhancing the learning experience of student radiographers with dyslexia. *Radiography*, 14(1), 32–38. <https://doi.org/10.1016/j.radi.2006.05.004>
- Frueh, S. (2023, November 6). How AI Is Shaping Scientific Discovery. *Nationalacademies.org*. <https://www.nationalacademies.org/news/2023/11/how-ai-is-shaping-scientific-discovery>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5(1). <https://doi.org/10.1186/1748-5908-5-69>
- Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, 2(6). <https://doi.org/10.1371/journal.pdig.0000278>
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024, March 4). GPT-4 technical report. *arXiv.org*. <https://doi.org/10.48550/arXiv.2303.08774>
- Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis*, 18(10), 2119–2126. <https://doi.org/10.11124/jbies-20-00167>
- Peters, M. D., Marnie, C., Colquhoun, H., Garritty, C. M., Hempel, S., Horsley, T., Langlois, E. V., Lillie, E., O'Brien, K. K., Tunçalp, Zge, Wilson, M. G., Zarin, W., & Tricco, A. C. (2021). Scoping reviews: Reinforcing and advancing the methodology and application. *Systematic Reviews*, 10(1). <https://doi.org/10.1186/>

- s13643-021-01821-3
- Qlik. (2024). How big data and Ai work together: Synergies & benefits. <https://www.qlik.com/us/augmented-analytics/big-data-ai>
- Russell, S. J., & Norvig, P. (2022). Artificial Intelligence: A modern approach Stuart J. Russell and Peter Norvig; *contributing writers: Ming-Wei Chang, Jacob Devlin, Anca Dragan*, (4th ed., Ser. global ed.). Pearson Education.
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). Prisma extension for scoping reviews (PRISMA-SCR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/m18-0850>
- Van Noorden, R., & Perkel, J. M. (2023). AI and science: What 1,600 researchers think. *Nature*, 621(7980), 672–675. <https://doi.org/10.1038/d41586-023-02980-0>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of Artificial Intelligence in achieving the Sustainable Development Goals. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-019-14108-y>
- Volino, L. R., Allen, S. M., & Gallimore, C. E. (2021). Addressing the challenges of providing accommodations for pharmacy students with disabilities across Learning Environments. *American Journal of Pharmaceutical Education*, 85(7), 8455. <https://doi.org/10.5688/ajpe8455>
- Wang, X., Huey, S. L., Sheng, R., Mehta, S., & Wang, F. (2024, April 21). SciDaSynth: Interactive structured knowledge extraction and synthesis from scientific literature with large language model. *arXiv.org*. <https://arxiv.org/abs/2404.13765>