

Using LLM (Large Language Model) to Improve Efficiency in Literature Review for Undergraduate Research

Shouvik Ahmed Antu¹, Haiyan Cheng² and Cindy K. Richards³

^{1,2,3} Willamette University, 900 State St, Salem, Oregon-97301

Abstract:

The potential of artificial intelligence (AI) to streamline and improve the research process for academics is becoming increasingly evident as this technology develops. A promising avenue for conducting literature reviews is to employ artificial intelligence (AI). Using OpenAI's ChatGPT, this paper explores the utility of this tool in the context of academic literature reviews. Our study focuses on how ChatGPT can be used to support the literature review process for undergraduate students conducting research for their capstone courses. Furthermore, we will explore the possible drawbacks and limitations of relying on artificial intelligence to perform such research tasks. The aim is to provide a balanced and comprehensive view of the role and future potential of AI, and specifically ChatGPT, in literature review studies. In addition to evaluating the accuracy and relevance of the results generated by ChatGPT, we will also examine the quality of results generated by ChatGPT.

Keywords

Large Language Model, ChatGPT, literature review, Undergraduate Research

Introduction

Many undergraduate programs include one capstone course to teach students how to conduct research. Because those research experiences not only help students develop their hands-on learning and problem solving skills, but also provide them with opportunities to explore their interests, and gain a deeper understanding of a specific area of study. In preparation for graduate school, undergraduate research experiences provide students with competitive advantages. In that class, students often learn about the foundations of the research process, e.g., reading a research paper, doing a literature review, identifying and communicating a research problem, etc. Among all the components of research, the literature review plays an important role in establishing the context for the study [11]. By doing so, students can establish the context of the research problem, build a theoretical framework, learn about relevant literature to avoid redundancy and replication. There is a steep learning curve for undergraduate students who have never done research before to conduct a literature review. Oftentimes, new undergraduate researchers struggle with identifying relative content, synthesizing information, understanding new content, figuring out the relationship of the resource, and writing in the genre of an academic literature review.

Recently, Artificial intelligence (AI) has emerged as a transformative technology, impacting various sectors of society from healthcare and finance to education and research [1, 2]. The development of sophisticated AI systems capable of understanding and generating human-like text has spurred significant changes in academic research. OpenAI's ChatGPT is one such advanced language model that redefines how we approach and conduct literature reviews [4].

¹AIED2023 Empowering Education with LLMs - the Next-Gen Interface and Content Generation
EMAIL: sahmed@willamette.edu (A. 1); hcheng@willamette.edu (A. 2); crichard@willamette.edu (A. 3)
ORCID: 0009-0005-0168-8577 (A. 1); 0000-0002-9711-5976 (A. 2);
© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In the current era of information overload, where new research is being published at an unprecedented rate, traditional approaches to literature review can be challenging and inefficient. This is where AI, and in particular, language models like ChatGPT, can play a useful role. With the help of ChatGPT, students can potentially carry out literature review more efficiently.

This paper aims to explore the potential of ChatGPT in literature review studies by providing an overview of the ChatGPT model, discusses its possible applications and benefits in conducting literature reviews, and offers insights into potential drawbacks and limitations.

ChatGPT: An Overview

The advent of ChatGPT from OpenAI denotes a remarkable advance in the world of artificial intelligence. This particular language model is a variation of the Generative Pretrained Transformer (GPT) models, which have been developed with the primary aim of producing text that closely resembles the manner in which human beings communicate, in response to specific prompts. The present approach leverages transformer neural networks, with specific emphasis on the transformer decoder variant, employing machine learning techniques to generate text that is both logically consistent and contextually relevant. One of the fundamental distinguishing factors of ChatGPT from its antecedent language models is characterized by the expansiveness and intricacy of its training data. The model has undergone training using heterogeneous internet texts that encompass a wide range of subjects, genres, and tonalities. This expansive set of data affords the artificial intelligence model with the capacity to effectively address a diverse range of tasks, encompassing intricate article synthesis to original content formation. ChatGPT uses unsupervised learning algorithms on unlabeled data. It acquires the ability to anticipate the succeeding word in a given sentence through scrutinizing the patterns present in the dataset on which it has been trained. This capability facilitates the generation of responses that bear a remarkable resemblance to those of human language, manifesting similarities particularly in regards to the textual organization and substance [5, 7].

It is important to acknowledge that even though ChatGPT exhibits aptitude in producing text that resembles that of a human being, it lacks the capacity for comprehension and consciousness. The responses produced by the subject under consideration stem exclusively from statistical pattern recognition. Despite the production of said responses, the subject lacks inherent comprehension of the meaning of the content generated. Hence, although the language model can mimic human conversation and produce a coherent essay, it lacks the cognitive ability to comprehend or develop judgments pertaining to the subjects it addresses. Nevertheless, there is continuing discourse concerning the ethical ramifications of such cutting-edge artificial intelligence technology, encompassing concerns related to confidentiality, potential misuse, and the consequences for the workforce.

Within the context of scholarly inquiry, and more specifically the synthesis and evaluation of existing literature, ChatGPT exhibits significant potential. The capacity to rapidly and precisely examine and condense substantial quantities of text could transform the procedure of literature review, enhancing its efficiency and comprehensiveness. In the sections that follow, we analyze this potential and the complexities that are associated with it.

ChatGPT for Literature Reviews

A literature review [10] is a fundamental part of research that entails the identification, collection, evaluation, and interpretation of relevant works on a specific topic. The traditional approach to literature reviews is a labor-intensive process that often requires weeks or months of work. Although there are some recommendations on successful literature review through a systematic approach [11], it still requires time consuming manual work. Researchers manually search databases, read through potentially relevant articles, and organize findings to synthesize a coherent overview of existing knowledge on a topic [12]. This laborious process poses significant challenges, including potential

oversight of relevant studies and biased selection of literature, leading to gaps in the review. To begin a literature survey, one needs to have a general direction, come up with relevant keywords. Those keywords can then be used to locate related literature. An experienced researcher who already possesses deep domain knowledge may be able to find the relevant resources quickly, but for students who have never done research, the first step is usually the most difficult. Considering the similarity of coming up with research keywords versus generating prompts for chatGPT, we propose using chatGPT to streamline literature reviews. In doing so, we can achieve the following:

Efficiency: The ability to process vast amounts of text quickly is one of ChatGPT's main strengths. In a literature review context, this could translate into sifting through hundreds or thousands of articles, abstracts, and research papers in a fraction of the time it would take a human researcher. Furthermore, ChatGPT can organize the retrieved literature, providing a structured overview that may speed up the review process.

Comprehensiveness: When tasked with conducting a literature review, ChatGPT's machine learning algorithms allow it to scan multiple databases and sources. This includes not just mainstream databases, but also more obscure or specialized ones. This ability to cover a wide breadth of sources increases the chances of a comprehensive and inclusive literature review.

Selection: Human researchers may unintentionally bring their biases into the selection process during literature reviews. However, as an AI, ChatGPT uses machine learning algorithms to select literature based on the specified criteria. It can be programmed to search for articles based on specific keywords, authors, journals, or a combination of these and other factors. ChatGPT asserts that its process can ensure that the literature review is exhaustive and unbiased.

Understanding and Summarization: A key part of any literature review is understanding and summarizing the collected articles. Despite lacking true understanding, ChatGPT can provide effective summarizations based on its pattern recognition capabilities. It can extract key points, conclusions, and methodologies from papers, thus offering a concise summary for each, which may save researchers time.

Gap Identification: Identifying gaps in existing literature is crucial for proposing new research avenues. With its ability to analyze large amounts of data, ChatGPT may help researchers identify trends and gaps in current research, guiding them towards areas that require further investigation.

Case Study: Application in an Undergraduate Research Capstone Class

We engaged ChatGPT to explore its capabilities to assist with literature reviews in undergraduate capstone courses. Specifically, we examined the utility of ChatGPT to assist with literature reviews in different academic disciplines: Computer Science and Communication Studies. Our multidisciplinary approach provides insight into how faculty and students in the sciences and humanities may effectively utilize ChatGPT to assist with literature reviews.

Initially, we prompted ChatGPT to “provide 3-4 examples of a suitable and compelling research topic within the humanities” in the context of a 4-credit undergraduate capstone research course for a Communication Studies major. ChatGPT generated four topics:

1. The Influence of Social Media on Body Image Perceptions and Mental Health
2. Digital Divide and its Impact on Remote Learning
3. The Role of Communication in Climate Change Advocacy
4. Misinformation and Fact-Checking in the Digital Age

These are relevant--but broad--areas of research for the discipline of Communication Studies. To examine the platform's capabilities to assist with the literature review process, we next prompted ChatGPT to “identify relevant literature” on each of these four research topics. ChatGPT responded with a bibliography of two peer reviewed journal articles per topic. Some of the articles ChatGPT identified are relevant and extensively cited. For example, when prompted to identify relevant

literature on the topic of the *Digital Divide and its Impact on Remote Learning*, ChatGPT highlighted “Reconsidering political and popular understandings of the digital divide” by Neil Selwyn, a 2004 article that is cited by more than 1800 publications. However, other results revealed weaknesses in ChatGPT. On the topic of the Digital Divide and its Impact on Remote Learning, ChatGPT cited a 2020 article entitled “Inequity of the digital divide: COVID-19 and online learning.” This article does not exist. Although this citation appeared realistic at first glance, when we worked with a research librarian to locate the article, it became clear that the source was the product of a ChatGPT hallucination. We did not identify hallucinations in the short bibliographies ChatGPT generated for other topics, but we did note that some of the sources ChatGPT suggested are not very relevant due to methodology or outdated findings. Excluding the 2020 citation hallucinated by ChatGPT, the sources it suggested for literature review did not include any work published after 2018.

After we reviewed these results we developed a new prompt, in an effort to elicit a more useful response from ChatGPT. We asked: “For each of these topics, identify relevant literature on this research topic. Focus on qualitative literature in the discipline of communication and identify at least 10 excellent sources.” In response ChatGPT produced a bibliography of three “seminal works in the field that would likely be relevant” to each research topic. As a whole, these results were more relevant and credible, although they were not as numerous as we requested. This time, for the topic *Digital Divide and its Impact on Remote Learning*, ChatGPT identified three existent sources:

- A. Warschauer, M. (2004). *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT Press;
- B. Van Dijk, J. A. (2006). Digital divide research, achievements and shortcomings. *Poetics*, 34(4-5), 221-235.
- C. Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New Media & Society*, 6(3), 341-362.

All three sources are influential; each one has been cited more than 1800 times. All three provide surveys of prior scholarship on the central topic of the “digital divide.”

Our experience illustrated the importance of human-in-the-loop, especially in the context of undergraduate research courses. If ChatGPT will be used to support student work on literature reviews, it is important for faculty and students to engage this tool with a strategic, critical process. For example, student researchers will likely need instructional support to enact a process that includes: developing effective prompts, critically examining ChatGPT responses, and using disciplinary knowledge to iterate.

As we enacted this process, we yielded more useful results from ChatGPT. For example, for the topic *Digital Divide and its Impact on Remote Learning*, we prompted ChatGPT to use the three sources it identified (Warschauer 2004, Van Dijk 2006, Selwyn 2004) in order to: "(1) Extract key points, and generate summaries of the literature, (2) Organize the retrieved literature and provide a structured overview, (3) "Use the retrieved literature to write 1500 words of in-depth analysis of existing works on the topic to provide context, identify gaps, and set the stage for further investigation." The responses ChatGPT generated were general, but the summations were accurate enough to serve as starting places for understanding the sources and key terms. The format of the “structured overview” and “1500 words of in-depth analysis” also modeled organizational structures that are common in academic writing; because of this the summations have potential to serve as scaffolds for students to build on, as they work to develop accurate, nuanced, and up-to-date literature reviews. Here again, our experience suggests that student researchers will likely need instructional support to enact a process that includes critically examining ChatGPT responses, using disciplinary knowledge to iterate. With a strategic approach, utilizing ChatGPT in addition to existing research processes and tools shows potential to help students identify potential starting places for literature review and adapt an organizational scaffold to support their insights.

While we see potential to strategically engage ChatGPT as a tool in undergraduate research courses, we also recognize a number of limitations of the current platform. ChatGPT asserts that it “offers immense potential in optimizing and streamlining literature reviews,” and that its strengths include providing “unbiased selection” of literature and “help[ing] researchers identify trends and gaps in current research, guiding them towards areas that require further investigation” (qtd. from above). However, ChatGPTs responses to our prompts raise some questions about literature selection and analysis. For example, for the topic *Digital Divide and its Impact on Remote Learning*, all of the

sources ChatGPT selected were published more than fifteen years ago, by scholars who appear to identify as white men, and have been cited more than 1500 times. When prompted to generate “in-depth analysis of existing works on the topic to provide context, identify gaps, and set the stage for further investigation,” ChatGPT wrote:

Although these works provide crucial insights into the digital divide, there is still an array of gaps that need addressing. The continuous evolution of technology demands continuous evolution in our understanding of the digital divide. Future research could benefit from examining the intersectionality of different factors contributing to the digital divide. This could include investigating the role of disability, gender, and race in digital inclusion and remote learning.

In this case, ChatGPT’s literature selection and summation did not include or acknowledge significant research that already exists on “the role of disability, gender, and race in digital inclusion and remote learning” or influential publications on the digital divide, digital inequality, and remote learning by scholars who are women, people of color, and/or people with disabilities (see for example Livingstone & Helsper 2007; Hargittai & Hinnant 2008; Jenkins, Ito & boyd 2015; Ali 2020). We cannot extrapolate larger patterns from this narrow test, but ChatGPT’s characterization of its process as “unbiased selection” obscures how this AI machine learning system amplifies some research findings and not others, and may not provide an “equal playing field” (Noble 2008). Because ChatGPT’s training data currently ends in 2021 we did not expect the platform to include the most contemporary literature on remote learning, but we were surprised to see that all the sources it selected on this topic predate 2008. These experiences illustrate that for faculty and student researchers who engage ChatGPT, it will be important to develop human-in-the-loop processes to identify and work beyond the platform’s elisions, in order to build research projects that engage the breadth, depth, and diversity of knowledge on the topic.

In a similar fashion, we asked chatGPT to “provide 3-4 examples of a suitable and compelling research topic in computer science”, and the following four topics are generated:

1. Ethical Considerations in Artificial Intelligence (AI) and Machine Learning (ML): Exploring the ethical implications of AI and ML algorithms, such as algorithmic bias, privacy concerns, and social impact.
2. Human-Computer Interaction (HCI) for Augmented and Virtual Reality (AR/VR): Investigating the design and usability aspects of AR/VR systems to enhance user experience and interaction.
3. Cybersecurity for Internet of Things (IoT) Networks: Examining security challenges and solutions for IoT devices and networks.
4. Natural Language Processing (NLP) for Text Summarization: Exploring techniques for automatic text summarization to extract key information from large volumes of text.

Most of the research topics are relevant and current, but they are in general very broad, and lacking individual niche, which could potentially lead to another survey style research paper as the research outcome instead of resulting scientific breakthroughs. We then asked chatGPT to “identify 2 most recent literature for each of the above 4 research topics”, and chatGPT returned the following result:

1. Ethical Considerations in Artificial Intelligence (AI) and Machine Learning (ML):
 - Paper: "Ethics of AI and Machine Learning Algorithms: A Critical Review" by Suresh and Gutta (2021)
 - This paper provides a critical review of the ethical considerations in AI and ML algorithms, discussing topics such as bias, fairness, interpretability, and accountability, and proposes guidelines for ethical algorithmic design.
 - Paper: "Towards Ethical Guidelines for Bias Mitigation in Machine Learning" by Madaio et al. (2020)
 - This study focuses on the ethical challenges related to bias in machine learning models. It proposes guidelines and techniques for bias detection and

- mitigation to ensure fairness and accountability in algorithmic decision-making.
- 2. Human-Computer Interaction (HCI) for Augmented and Virtual Reality (AR/VR):
 - Paper: "Towards Natural and Intuitive Interaction in Augmented Reality" by Bruder et al. (2021)
 - This paper discusses natural and intuitive interaction techniques for augmented reality systems. It explores gesture recognition, spatial mapping, and haptic feedback to enhance user experiences in AR applications.
 - Paper: "Virtual Reality for Psychological Therapy: Recent Advances and Future Directions" by Valmaggia et al. (2020)
 - This review paper explores the use of virtual reality in psychological therapy. It discusses recent advances in VR-based interventions, user experiences, and challenges for incorporating VR technologies into therapeutic practices.
- 3. Cybersecurity for Internet of Things (IoT) Networks:
 - Paper: "A Survey on Security and Privacy Issues in IoT-Based Healthcare Systems" by Malik et al. (2021)
 - This survey paper focuses on security and privacy challenges in IoT-based healthcare systems. It discusses various threats, vulnerabilities, and countermeasures to protect sensitive healthcare data in IoT networks.
 - Paper: "A Lightweight Blockchain-Based Authentication Protocol for IoT Devices" by Alrawais et al. (2020)
 - This paper proposes a lightweight blockchain-based authentication protocol for securing IoT devices. It addresses security and scalability challenges in IoT networks by leveraging blockchain technology.
- 4. Natural Language Processing (NLP) for Text Summarization:
 - Paper: "Pretrained Transformers Improve Out-of-Distribution Robustness" by Wang et al. (2021)
 - This paper explores the use of pre trained transformer models in text summarization tasks, discussing their impact on out-of-distribution robustness and generalization capabilities.
 - Paper: "Extractive Summarization of Long Documents by Combining Global and Local Context" by Liu et al. (2020)
 - This research focuses on extractive summarization of long documents. It proposes a method that combines global and local context information to generate more informative and concise summaries.

At the first glance, the returned literature are mostly review papers themselves and indeed closely follow the research topic, which is a common drawback of using chatGPT—overgeneralization. After careful examination, as we try to find the actual papers, we discovered that the suggested “literature” does not exist most of the time. This creates a major problem of research credibility as citation often needs to be rigorous and uniform. What we need at this stage is to introduce human-in-the loop chatGPT, to follow the lead of chatGPT and further refine the topics and locate the exact paper. For example, by browsing through the summary returned with the first ethical paper, we immediately identify more keywords: “bias”, “fairness”, “interpretability”, “accountability”, “ethical algorithm design”, which prompt us to further narrow down the topic. When returning that literature, chatGPT did return a phrase “Please note that the availability and citation counts of research papers may vary, and it's important to conduct a comprehensive search using academic databases to find the most recent and highly cited publications in the respective research areas.” So we improved the prompt by asking chatGPT to “perform a comprehensive search using academic databases to find the most recent and highly cited 2 publications for each of the 4 research areas”. After a couple of “apologies” from chatGPT, we got some guidelines. Here is an example for the first topic:

1. Ethical Considerations in Artificial Intelligence (AI) and Machine Learning (ML):
 - Start by searching academic databases like Google Scholar or IEEE Xplore using keywords such as "AI ethics," "machine learning ethics," or specific ethical concerns in AI/ML.

- Refine your search results by selecting a specific time frame for publication, such as the last 5 years.
- Sort the results by citation count to identify highly cited papers.
- Read the abstracts, introductions, and conclusions of the selected papers to determine their relevance and citation impact.

which loops back to the traditional ways of literature review.

If we ask chatGPT these two questions again, another four different topics will be generated. For one thing, we think this is definitely a great way to inspire students to think along a general direction, and further their inquiries that lead to better research topics.

Our case studies show that although chatGPT can help with the literature review process, it is crucial to understand that the final responsibility for the review's quality and integrity still lies with the human researcher. For instance, understanding the broader context, making connections across disciplines, and identifying potential bias in the studies reviewed require human cognitive abilities and expertise that AI currently lacks. It is essential to reflect on the possible downsides and constraints associated with this approach while conducting literature evaluations.

Quality Control: The vast computing abilities possessed by ChatGPT could lead to an overwhelming overflow of information, some of which may not be relevant or accurate. Even with advanced algorithms, the artificial intelligence model may give inaccurate summaries or analysis due to its tendency to misinterpret intricate concepts or subtleties found in textual data.

Contextual Understanding: As an artificial intelligence, ChatGPT is deficient in genuine comprehension of the material it analyzes. Although proficient in recognizing patterns and producing coherent written content, it lacks the capacity to comprehend the meaning or significance of the conveyed information. Within the context of literature reviews, the requirement for comprehensive comprehension of the subject matter frequently proves to be a noteworthy constraint.

Reliance on Training Data: The performance of ChatGPT is primarily contingent upon the caliber and multiplicity of its training data. Should the training data exhibit biases or limitations in its scope, the potential exists for an adverse impact on the ability of the AI system to access a full spectrum of literature.

Interdisciplinary research necessitating diverse knowledge and expertise may not be optimally facilitated by ChatGPT, owing to its limited comprehension of the holistic nature of such inquiries.

Ethical Concerns: The utilization of artificial intelligence (AI) tools, such as ChatGPT, in academic research raises numerous ethical concerns. An exemplification of critical importance is the assurance of the non-compromise of confidentiality and privacy of the data utilized for the training of the model, in the context of implementing AI. It is crucial for researchers to maintain vigilance regarding the capacity of artificial intelligence to be employed in generating or circulating fraudulent or discriminatory research investigations.

Accessibility and Costs: As the utilization of artificial intelligence (AI) tools continues to expand, it is possible that not all individuals will have access to sophisticated models, such as ChatGPT-4, due to factors such as financial limitations or technical prerequisites. The aforementioned scenario has the potential to result in an unequal allocation of resources within the research community.

While these drawbacks present challenges, they also offer avenues for improvement. By acknowledging these limitations, researchers can work towards refining the AI's application in literature reviews, ensuring rigorous supervision, cross-verification of information, and using AI as a supplement to, rather than a replacement for, human expertise and judgment.

Conclusion

In conclusion, leveraging AI tools such as ChatGPT offers the potential to support academic literature reviews, providing efficiency and comprehensive coverage while eliminating human biases. While the application of AI in literature reviews is still in its infancy, and challenges remain, the benefits are promising. It's time to consider AI not as a replacement for human researchers but as a powerful tool that can enhance the research process. AI can help researchers stay organized and find new connections between their sources. It can also help them identify gaps in their research that need to be filled.

References

- [1] Sabzalieva, Emma, and Arianna Valentini. "ChatGPT and artificial intelligence in higher education: quick start guide." (2023).
- [2] Holmes, Wayne, Maya Bialik, and Charles Fadel. "Artificial intelligence in education." Globethics Publications, (2023) 621-653.
- [3] Qureshi, Basit. "Exploring the use of chatgpt as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges." arXiv preprint arXiv:2304.11214 (2023).
- [4] Kasneci, Enkelejda, et al. "ChatGPT for good? On opportunities and challenges of large language models for education." *Learning and Individual Differences* 103 (2023).
- [5] Hariri, Walid. "Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing." arXiv preprint learning, commerce, and politics. John Wiley & Sons.
- [6] Noble, S.U., 2018. Algorithms of oppression. In *Algorithms of oppression*. New York University Press. arXiv:2304.02017 (2023).
- [7] Farrokhnia, Mohammadreza, et al. "A SWOT analysis of ChatGPT: Implications for educational practice and research." *Innovations in Education and Teaching International* (2023): 1-15.
- [8] Ali, Wahab. "Online and remote learning in higher education institutes: A necessity in light of COVID-19 pandemic." *Higher education studies* 10.3 (2020): 16-25.
- [9] Jenkins, Henry, and Mizuko Ito. *Participatory culture in a networked era: A conversation on youth, learning, commerce, and politics*. John Wiley & Sons, 2015.
- [10] Papaioannou, Diana, Anthea Sutton, and Andrew Booth. "Systematic approaches to a successful literature review." *Systematic approaches to a successful literature review* (2016): 1-336.
- [11] Boote, David N., and Penny Beile. "Scholars before researchers: On the centrality of the dissertation literature review in research preparation." *Educational researcher* 34.6 (2005): 3-15.
- [12] Frank, Hermann, and Isabella Hatak. "Doing a research literature review." *How to get published in the best entrepreneurship journals* (2014): 94-117.