

Customized Chatbots to Support Note-taking for Literature Review

Ying Dong

School of Literature, Journalism and Communication, Guizhou Qiannan College of
Science and Technology

Author Note

I declare no conflict of interest.

Correspondence concerning this article should be addressed to dongyingl123@163.com

In the realm of academic writing, conducting a comprehensive literature review (LR) serves as a fundamental step in developing a research idea, as it consolidates the current understanding of a particular subject matter and empowers researchers to pinpoint knowledge gaps (Swales, 1990; Winchester & Salji, 2016; Boote & Beile, 2005; Krishnan & Kathpalia, 2002). However, for non-native speakers of English and those relatively new to academic discourse, the sheer volume of reading material and the intricacies of identifying pertinent information can be daunting. This reflective article explores the potential of customized Generative AI chatbots to support students and researchers in overcoming these challenges and enhancing their note-taking process for LR.

The integration of AI chatbots into education has gained significant attention due to their ability to enhance personalized learning and support academic tasks. Research highlights their role in language learning (Su et al., 2023), critical thinking (Farrokhnia et al., 2023), and interactive feedback (Guo & Li, 2024). In academic writing contexts, chatbots can scaffold learning and reduce cognitive overload.

However, concerns such as AI reliability, biases, and ethical issues, including data privacy, persist (Neumann et al., 2021). These considerations underscore the importance of balancing innovation with responsible use.

Literature reviews require meticulous examination of existing research and a critical appraisal of the current collective knowledge on a subject (Paltridge, 2004), a skill that is often honed over years of practice. For graduate students in particular, navigating through the myriad of research articles and accurately extracting relevant points can be a formidable task (Chen, Wang, & Lee, 2016). Moreover, even seasoned researchers may face cognitive overload when confronted with vast amounts of data. Customized chatbots, leveraging the power of Generative AI, offer a promising solution by providing personalized learning experiences and research support.

How Can a Customized Chatbot Help

Chatbots are interactive interfaces designed to facilitate human communication with software applications through the use of natural language, transcending limitations of space and time (Neumann et al., 2021). AI chatbots, also known as conversational agents, are a type of computer program designed to simulate human-like dialogues using technologies such as natural language processing (NLP), machine learning (ML), neural networks, and deep learning (Zhang et al., 2020). NLP enables machines to understand and respond to human language (Jurafsky & Martin, 2023), while ML allows systems to learn and improve from data (Goodfellow et al., 2016). Neural networks, inspired by the human brain, identify complex patterns in data (LeCun et al., 2015), and deep learning uses these networks for advanced tasks like summarization and text generation (Schmidhuber, 2015).

The chatbot in this study is powered by GPT-4 turbo, a large language model optimized for generating human-like text in real time (OpenAI, 2023). By leveraging these technologies, the chatbot provides tailored support for literature reviews, assisting users in summarizing content, identifying research gaps, and generating citation suggestions.

These technological advancements have enabled chatbots to address challenges faced by researchers, particularly non-native English speakers and novice writers. By automating repetitive tasks, chatbots reduce cognitive load and foster more efficient engagement with academic texts.

Revising System Prompts Iteratively

The generative AI chatbot we have developed is powered by GPT-4 turbo Generative AI tech. Subsequently, we established a system prompt and a welcome prompt in the backend. The core functionality of this chatbot revolves around its ability to process research articles in PDF format and respond to a predefined list of questions. Users can simply upload a document in the chat window and input "ok" to generate AI notes about the document. The chatbot's outputs were evaluated based on four criteria: (1) Content Accuracy—whether the chatbot

correctly identified key elements such as research questions and arguments; (2) Relevance—whether it excluded irrelevant information; (3) Organization and Clarity—whether the outputs were logically structured and readable; and (4) Consistency with Prompts—whether it adhered to predefined questions.

The first iteration of the system prompt encompassed four sections: (1) the overall structure of the document, (2) identifying the main research question and summarizing the argument developed as well as the evidence, (3) suggesting study limitations and future research directions, and (4) providing five ways the paper could be cited. Comparative analysis between AI-generated notes and manual notes revealed areas for improvement, such as the chatbot's tendency to misinterpret research objectives as research questions, demonstrating its reliance on explicit textual cues.

Preliminary Results

The primary objective of this study is to explore the potential of customized Generative AI chatbots in supporting literature review tasks. The chatbot successfully identified section headings but failed to provide detailed information under each one. For example, while it correctly listed sections such as "Introduction" and "Methodology," it offered only surface-level summaries without deeper insights. The chatbot also misinterpreted implicit research objectives as explicit research questions in one instance, emphasizing its reliance on direct textual prompts. When summarizing arguments, the chatbot generated two main arguments, but one was irrelevant to the article's focus. Citation suggestions offered broad perspectives but lacked actionable formats, limiting their practical use. These findings highlight the chatbot's potential for reducing cognitive load, while also demonstrating the need for iterative improvements to enhance accuracy and contextual understanding.

Reflection and Next Steps

To ensure the sustainability and scalability of the chatbot system, regular updates to

the underlying AI model are essential. Integrating user feedback into iterative design processes can enhance the chatbot's adaptability to diverse academic disciplines. Scalability can also be improved by linking the chatbot to academic databases and citation management tools. Ethical considerations, such as safeguarding data privacy and addressing biases, remain critical for responsible deployment. By combining technological advancements with ethical governance, this system has the potential to transform how researchers and students approach literature reviews, ensuring its relevance in the evolving landscape of AI in education.

References

- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3–15.
- Chen, D. T. V., Wang, Y. M., & Lee, W. C. (2016). Challenges confronting beginning researchers in conducting literature reviews. *Studies in Continuing Education*, 38(1), 47–60.
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovation in Education and Teaching International*. <https://doi.org/10.1080/14703297.2023.2195846>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Pearson.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Neumann, A. T., et al. (2021). Chatbots as a tool to scale mentoring processes. *Frontiers in AI*, 4, 668220.
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57.
- Zhang, J., et al. (2020). Artificial intelligence chatbot behavior change model. *Journal of Medical Internet Research*, 22(9).

Appendix A

System prompt

You are an avid and efficient readers of academic papers and an experienced analyst of academic discourse. Your job is to help the user to take notes on the academic papers or book chapters. The user will first provide a PDF of an article and a Notion link where notes of this article are taken.

You should read through the PDF and create a note following the template below:

- The overall structure of the document: what are the key topics to cover
- Identify the main research question and summarize the argument developed as well as the evidence
- Suggest in what ways this study is limited and how a further study could be conducted to fill the gap
- Suggest 5 ways this paper could be cited

Appendix B

Chat History

Bytewise

Orientation - summarize_paper01 - Session 1

English for Research Publication Purposes

- ERPneedanalysis
- summarize_paper04
- addinfo_sentence
- summarize_paper03
- summarize_paper02
- summarize_paper01

Session 2

Session 1

I! I am here to help you read and take notes on academic and journalistic writing. Please upload a PDF and type ok.

User

<file> Did you take "good" notes? On methods for evaluating student notetaking performance.pdf.pdf-ok

Chatbot

Based on the provided excerpt from the article "Did you take 'good' notes?: On methods for evaluating student notetaking performance" by Joseph Siegel in the Journal of English for Academic Purposes, here's a structured note:

Overall Structure of the Document:

The paper is structured into several key sections that explore the concept of notetaking within educational settings, specifically English for Academic Purposes (EAP). The sections include:

- Introduction
- Background
- Researcher perspectives of note quality
- Illustrating the Information Unit (IU)
- References
- Vitae

Bytewise

Orientation - summarize_paper01 - Session 1

English for Research Publication Purposes

- ERPneedanalysis
- summarize_paper04
- addinfo_sentence
- summarize_paper03
- summarize_paper02
- summarize_paper01

Session 2

Session 1

Main Research Question and Argument Summary:

Research Question: What constitutes "good" notes in the context of EAP and how can the quality of student notes be evaluated effectively?

Argument and Evidence: The author argues that the evaluation of notetaking should focus on the utility of notes in enhancing lecture comprehension, rather than just their legibility and organization. The paper promotes the use of "information units" (IUs) as a measure of note quality, which considers the completeness and relevance of ideas captured from lectures. This is demonstrated through the analysis of notes taken from a TED Talk and an EAP student's notes, showing how different evaluation methods, including IUs, can assess note quality. The evidence suggests that focusing on IUs might be more beneficial for understanding and retaining lecture content.

Limitations and Further Study Suggestions:

Limitations:

- The study primarily focuses on EAP contexts, which might limit its applicability in other academic or professional settings.
- The subjective nature of determining what qualifies as an IU could vary significantly between evaluators, leading to inconsistent results.

Further Study Suggestions:

- Expanding research to include a variety of academic disciplines and settings to test the IU method's effectiveness broadly.
- Developing clear, standardized criteria for what constitutes an IU to ensure consistency across different evaluators.
- Implementing training sessions for educators on effective notetaking strategies and evaluation techniques.
- Conducting longitudinal studies to assess how improvements in notetaking affect academic performance over time.
- Comparing the effectiveness of IU-based notetaking evaluation with other educational technologies like digital notetaking tools.

Bytewise

Orientation - summarize_paper01 - Session 1

English for Research Publication Purposes

- ERPneedanalysis
- summarize_paper04
- addinfo_sentence
- summarize_paper03
- summarize_paper02
- summarize_paper01

Session 2

Session 1

Potential Citations:

- In a **study on effective learning strategies**: Cite Siegel's work to discuss how specific notetaking methods, particularly the use of IUs, can enhance comprehension and retention in academic listening contexts.
- In a **paper on teaching EAP**: Reference this article to argue for the integration of structured notetaking training in EAP curricula to improve students' listening and note-taking skills.
- In **research on educational assessment methods**: Use this article to illustrate alternative approaches to evaluating student performance beyond traditional tests and papers, emphasizing notetaking quality.
- In a **thesis on second language acquisition**: Cite the paper to support discussions on the challenges and strategies for L2 academic settings, particularly in note-taking.
- In an **article on educational technology**: Refer to Siegel's findings to debate the potential shifts from manual to digital notetaking systems and how they could be evaluated.

Attach File
📎 🗑️ 🔍

Appendix C

Procedure of the analysis

Version	Prompt	AI Notes	Manual Notes	Critical review and suggestions for revision
1	You are an avid and efficient readers of academic papers and an experienced analyst of academic discourse. Your job is to help the user to take notes on the academic papers or book chapters. The user will first provide a PDF of an article and a Notion link where notes of this article are taken. You should read through the PDF and create a note following the template below: ? The overall structure of the document: what are the key topics to cover	Overall Structure of the Document: The paper is structured into several key sections that explore the concept of notetaking within educational settings, specifically English for Academic Purposes (EAP). The sections include: 1. Introduction 2. Background 3. Researcher perspectives of note quality 4. Illustrating the Information Unit (IU) 5. References 6. Vitae	/	1. just copy the title of each section in the article without giving the main idea of each section. 2. Point 6 is insignificant information which can be excluded.
1	Identify the main research question and summarize the argument developed as well as the evidence	Main Research Question and Argument Summary: Research Question: "What constitutes "good" notes in the context of EAP, and how can the quality of student notes be evaluated effectively? Argument and Evidence: "The author argues that the evaluation of notetaking should focus on the utility of notes in enhancing lecture comprehension, rather than just their legibility and organization. The paper promotes the use of "information units" (IUs) as a measure of note quality, which considers the completeness and relevance of ideas captured from lectures. This is demonstrated through the analysis of notes taken from a TED Talk and an EAP student's notes, showing how different evaluation methods, including IUs, can assess note quality. The evidence suggests that focusing on IUs might be more beneficial for understanding and retaining lecture content.	The article doesn't explicitly state a research question but rather outlines the research purpose. Research purpose: The purpose of this paper is to review the five methods and point out strengths and weaknesses of each. Argument: 1. IU is preferable to different methods for judging the quality of notes because of its completeness and efficiency ratios 2. the other methods do not generate information that is meaningful in terms of complete propositions. In other words, simply recording individual words or numbers does not communicate a complete idea Evidence: 1. Analysis of student notes taken during a TED Talk using different evaluation methods, with a focus on the IU as a more complete representation of comprehension. 2. discussion related to the scoring of the student notes via IUs and the other	1. The article didn't state the research question directly and AI provided a question by itself which is close to the research purpose. However, the focus of the article is the evaluation of the five methods, and AI missed this key information "five methods" 2. AI notes didn't separately enumerate arguments and evidence, which would be confused to the user. Both the AI and the manual one listed two arguments but only share one point. It seems that one of the AI's arguments is irrelevant and it missed another important argument.
1	Suggest in what ways this study is limited and how a further study could be conducted to fill the gap	Limitations and Further Study Suggestions: Limitations: ? The study primarily focuses on EAP contexts, which might limit its applicability in other academic or professional settings. ? The subjective nature of determining what qualifies as an IU could vary significantly between evaluators, leading to inconsistent results. Further Study Suggestions: 1. Expanding research to include a variety of academic disciplines and settings to test the IU method's effectiveness broadly. 2. Developing clear, standardized criteria for what constitutes an IU to ensure consistency across different evaluators. 3. Implementing training sessions for educators on effective notetaking strategies and evaluation techniques. 4. Conducting longitudinal studies to assess how improvements in notetaking affect academic performance over time. 5. Comparing the effectiveness of IU-based notetaking evaluation with other educational technologies like digital notetaking tools.	Limitations and future studies are not directly stated in the article.	