

# Review of: "OpenAI ChatGPT Generated Content and Similarity Index: A study of selected terms from the Library & Information Science (LIS)"

Amit Verma

Potential competing interests: No potential competing interests to declare.

Plagiarism can be an issue with any language generation model, including GPT models. Since GPT models are trained on a vast amount of text data, they can generate human-like responses that may resemble existing content or be similar to previously published works.

In the current study, the authors have used similarity checking using Turnitin software but the topics that are listed - "citation in research", "information retrieval", "research data management" are very generic and the GPT responses may have overlapped with data sources that can potentially be outside of scope for Turnitin and hence difficult to validate if it has been plagiarized. The authors should expand the scope of these terms to include rare terms/non-generic phrases with limited literature available to test both the accuracy as well as plagiarism.

Furthermore, while GPT models while may generate coherent statements, they may not always be accurate. It may be important to address the issue of accuracy in the article and highlight some of the limitations of such models as an be seen in the following article:

<https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/?sh=5d9aa9447c7f>