2-22-2023

Generative AI as a Tool for Environmental Health Research Translation

Lauren B. Anderson
University of Louisville

Dhiraj Kanneganti

Mary Bentley Houk

Rochelle H. Holm
University of Louisville, rochelle.holm@louisville.edu

Ted Smith

Generative AI as a Tool for Environmental Health Research Translation

Lauren B. Anderson[1,2] [ORCID], **Dhiraj Kanneganti[1], Mary Bentley Houk[1], Rochelle H. Holm[1]** [ORCID], **and Ted Smith[1,3]** [ORCID]

[1]Christina Lee Brown Envirome Institute, School of Medicine, University of Louisville, Louisville, KY, USA, [2]Department of Urban and Public Affairs, College of Arts and Sciences, University of Louisville, Louisville, KY, USA, [3]University of Louisville Superfund Research Center, Louisville, KY, USA

Abstract One valuable application for generative artificial intelligence (AI) is summarizing research studies for non-academic readers. We submitted five articles to Chat Generative Pre-trained Transformer (ChatGPT) for summarization, and asked the article's author to rate the summaries. Higher ratings were assigned to more insight-oriented activities, such as the production of eighth-grade reading level summaries, and summaries highlighting the most important findings and real-world applications. The general summary request was rated lower. For the field of environmental health science, no-cost AI technology such as ChatGPT holds the promise to improve research translation, but it must continue to be improved (or improve itself) from its current capability.

Plain Language Summary This study explored the use of generative artificial intelligence (AI), specifically Chat Generative Pre-trained Transformer (ChatGPT), to summarize environmental health research articles. The field of environmental health sciences exemplifies this opportunity given the specialized language surrounding environmental contamination. Four differerent ChatGPT-generated summaries were evaluated from each of five articles. The average rating of summaries indicated good content quality, though sometimes removed important details or had minor inaccuracies. This study suggests that no-cost AI technology such as ChatGPT holds the promise to improve research translation to support environmental justice communities, mainstream media outlets, and community science groups, but within some boundaries.

1. Introduction

Generative artificial intelligence (AI), popularized by services like Chat Generative Pre-trained Transformer (ChatGPT), has been the source of much recent popular attention for publishing health research (Liebrenz et al., 2023; van Dis et al., 2023). The software can instantly create sophisticated content ranging from text-to-image to text-to-audio and video depending on user prompts (Gozalo-Brizuela & Garrido-Merchan, 2023). ChatGPT, released by OpenAI (2023; San Francisco, CA, USA) in November of 2022, is a large language model (LLM) trained on a vast amount of data that generates human-like responses to queries. One potentially valuable application is the summarization of studies in support of research translation intended for non-specialist readers. The field of environmental health sciences exemplifies this opportunity given the specialized language surrounding environmental contamination which often relies on knowledge of chemistry, toxicology, epidemiology, and physical sciences. Very often, concerned members of the public are excluded from that primary literature due to these barriers. These might include environmental justice communities, mainstream media outlets, and community science groups. Plain language summaries (PLSs) for a nonexpert audience should improve on accessibility, understanding, knowledge, communication of research, and empowerment (Stoll et al., 2022). As a complement, research translation activities should include the identification of key messages and development of messages that are easily understood across a variety of audiences (Grimshaw et al., 2012). Given the number of challenges with writing peer-reviewed research and the follow-on cost and effort in additionally creating PLSs, generative AI may offer a valuable new tool for expanding dissemination of environmental health risk research and providing access to broader constituencies.

LLM generative AI such as ChatGPT could greatly advance environmental health research by making an important contribution to wider dissemination of PLSs to non-experts who are often important in regulatory and other policy development. This should not be done uncritically but by vigilantly weighing the strengths and weaknesses

of these applications of generative AI. On one hand, it could enhance fairness, diversity, and accessibility in science (van Dis et al., 2023). On the other hand, it may pose risks by undermining research quality and trust. Chatbots can produce convincing but inaccurate text, which might undermine the very trust that research summarization aims to build (Liebrenz et al., 2023; van Dis et al., 2023). The role of academic researchers in environmental health programs is not limited to methods, but also the essential working relationships across audiences for translation to public health action (Hoar et al., 2022). We explored the use of ChatGPT as an alternative to traditional human expert written summarization to support PLSs as a form of research translation and provide suggestions for environmental health research summarization.

2. Methods

A convenience sample of five peer-reviewed environmental health articles were used to generate four different summaries using generative AI (OpenAI, 2023). The inclusion criteria for the selected manuscripts were: recently published (2021–2022), open-access, peer-reviewed published articles, and authored by colleagues who were University of Louisville environmental health investigators and collaborators with relevance to the general public (Supporting Information S1). The PLSs were created by entering the full text of each article into the ChatGPT interface with the prompt: "read the following," followed by a series of prompts: "Summarize the findings of this study in 500 words," "Summarize the research paper at an eighth-grade reading level," "What is the most important finding of this study?" and "What are the real-world impacts of this study?." These prompts were selected because they reflect the effort to identify each study's key messages for research translation purposes (Grimshaw et al., 2012), the intended audience of the ChatGPT-generated summaries was a non-expert. The level of an eighth grade readability was chosen as a metric of comprehension comparable to a typical community newspaper. Study authors who were either first, senior, or corresponding were contacted to review the summarizations. The author was provided rating instructions (Supporting Information S1) and rated each summary one time. The responses from ChatGPT were evaluated using a combination of Likert-scale, yes/no, and text to assess scientific accuracy, completeness, and readability at an eighth grade level by one study author. The author reviewers were blinded to the nature of the summarization; instructions stated that "summaries were written about their work." The use of generative AI was not disclosed nor was brought up to the authors conducting the ratings. Due to design limitations of using an author of each paper as rater, inter-rater reliability could not be assessed.

3. Results and Discussion

The average rating of summaries across the five studies ranged between 3 and 5 (average rating of 3.9). A score of 1 indicated poor quality and where 5 indicated good overall content quality. ChatGPT's general summary request was consistently rated lower than the other summary types (average rating of 3.4). Higher ratings were assigned to the more synthetic, insight-oriented activities, such as identifying the most important finding (average rating of 4.1), the production of a PLS suitable for an eighth grade reading level (average rating of 3.9), and real-world research applications (average rating of 3.9). These more insight-orientated summaries were also judged acceptable for use with the public. Even across this limited set of articles, two authors commented that language used by the ChatGPT-generated PLS was still too technical for general audiences. For instance, ChatGPT-generated content stated "controlled for the potential confounding effects of race and wealth" or that methods using "geographically clustered probability sampling design is representative of the US civilian non-institutionalized population." However, in some cases, ChatGPT's attempt at simplification removed important detail—for instance, stating wellbeing measures were associated with cardiovascular disease, rather than risk of cardiovascular disease. Minor inaccuracies around study method interpretations were also observed. In one article the ChatGPT-generated PLS referenced questions about levels of pollution and greenness around the home when the study did not include these questions. This finding on inaccuracies is similar to medical discharge summaries created with automation which have been shown to still require manual checking by a human expert since an inaccuracy in a summary report may lead to patient safety issues (Patel & Lam, 2023). The current discourse in scientific writing presents concerns about the potential for AI to generate human-like responses, thus bringing up the need for writing source disclosure (Gao et al., 2023).

There is a cost (economic, effort and time) of producing PLSs, for example, a Cochrane PLS is reported to take about 5 hr to write (Pitcher et al., 2022). In contrast, the no-cost ChatGPT summary requests in this study took only seconds to generate. In an age of misinformation (Spitale et al., 2023), ChatGPT-generated summaries may

also disincentivize non-experts to understand environmental health data. Errors such as disease association rather than risk-for-disease association could be quite troublesome.

In academic settings, while ChatGPT has been listed as a co-author (Frye, 2023) it has also been argued that this should not apply because AI cannot be held accountable for scientific writing (van Dis et al., 2023) or contacted for further information to understand environmental health data. There also needs to be discussion around the notion of copyright and implications for Fair Use laws. The publishers own the copyright to these articles, and unless they are open access there is a potential gray area with employing ChatGPT (itself from a private company) to generate new content from this existing and copyrighted content even for research translation use.

4. Conclusion

There are several critical, yet under-delivered, research needs, especially around inclusion and environmental health justice research (van Dis et al., 2023). This is a case where AI might do more good than harm and help level the playing field, for example, by creating accessible insights and enabling the large-scale production of high-quality PLSs which could improve open access to scientific information. This possibility, combined with the increasing public policy trends encouraging open access for research supported with public funds, may alter the role journal publications play in communicating science in society. Recommendations for future research include adding ratings by community members of paired (human vs. AI) PLSs; increasing the number and variety of environmental health research study types with more reviewers; setting safeguards to guard against the spread of misinformation; determining whether AI PLSs are culturally appropriate; and studying the efficiency of the AI PLS method. Having ratings done by an author of the article is suited to evaluating accuracy but may not be the best choice for other characteristics. The authors of this study did not consider the potential interpretability from a diverse set of audiences including communities that may be impacted by the research more than others. While not within the scope of this review, this would be an important caution and future direction to consider when analyzing the utility of ChatGPT for advancing translation of research for a broader lay audience. For the field of environmental health science, no-cost AI technology such as ChatGPT holds the promise to improve research translation, but it must continue to be improved (or improve itself) from its current capability.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The list of peer-reviewed articles analyzed during the current study are available in the supplement to this manuscript.

References

Frye, B. L. (2023). Should using an AI text generator to produce academic writing be plagiarism? *33 Fordham Intellectual Property, Media & Entertainment Law Journal*, *947*, 23.

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine*, *6*(1), 75. https://doi.org/10.1038/s41746-023-00819-6

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A state of the art review of large generative AI models. arXiv, 2301.04655. https://doi.org/10.48550/arXiv.2301.04655

Grimshaw, J. M., Eccles, M. P., Lavis, J. N., Hill, S. J., & Squires, J. E. (2012). Knowledge translation of research findings. *Implementation Science*, *7*(1), 1–17. https://doi.org/10.1186/1748-5908-7-50

Hoar, C., McClary-Gutierrez, J., Wolfe, M. K., Bivins, A., Bibby, K., Silverman, A. I., & McLellan, S. L. (2022). Looking forward: The role of academic researchers in building sustainable wastewater surveillance programs. *Environmental Health Perspectives*, *130*(12), 125002. https://doi.org/10.1289/EHP11519

Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D., & Smith, A. (2023). Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. *The Lancet Digital Health*, *5*(3), e105–e106. https://doi.org/10.1016/S2589-7500(23)00019-5

OpenAI. (2023). ChatGPT (version 3.5). Retrieved from https://openai.com

Patel, S. B., & Lam, K. (2023). ChatGPT: The future of discharge summaries? *The Lancet Digital Health*, *5*(3), e107–e108. https://doi.org/10.1016/S2589-7500(23)00021-3

Pitcher, N., Mitchell, D., & Hughes, C. (2022). Template and guidance for writing a Cochrane Plain language summary. Retrieved from https://training.cochrane.org/guidance-writing-cochrane-plain-language-summary.pdf

Spitale, G., Biller-Andorno, N., & Germani, F. (2023). AI model GPT-3 (dis) informs us better than humans. *Science Advances*, *9*(26), eadh1850. https://doi.org/10.1126/sciadv.adh1850

Stoll, M., Kerwer, M., Lieb, K., & Chasiotis, A. (2022). Plain language summaries: A systematic review of theory, guidelines and empirical research. *PLoS One*, *17*(6), e0268789. https://doi.org/10.1371/journal.pone.0268789

van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, *614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

References From the Supporting Information

Coleman, C. J., Yeager, R. A., Pond, Z. A., Riggs, D. W., Bhatnagar, A., & Pope, III, C. A. (2022). Mortality risk associated with greenness, air pollution, and physical activity in a representative US cohort. *Science of the Total Environment*, *824*, 153848. https://doi.org/10.1016/j.scitotenv.2022.153848

Coleman, C. J., Yeager, R. A., Riggs, D. W., Coleman, N. C., Garcia, G. R., Bhatnagar, A., & Pope, C. A. (2021). Greenness, air pollution, and mortality risk: A US cohort study of cancer patients and survivors. *Environment International*, *157*, 106797. https://doi.org/10.1016/j.envint.2021.106797

El-Mallakh, T. V., Hedges, S., Rai, J. P., Bhatnagar, A., Moyer, S., & El-Mallakh, R. S. (2022). Suicide and homicide more common with limited urban tree canopy cover. *Cities and the Environment (CATE)*, *14*(2), 4. https://doi.org/10.15365/cate.2022.140204

McLeish, A. C., Smith, T., Riggs, D. W., Hart, J. L., Walker, K. L., Keith, R. J., et al. (2022). Community-based evaluation of the associations between well-being and cardiovascular disease risk. *Journal of the American Heart Association*, *11*(22), e027095. https://doi.org/10.1161/JAHA.122.027095

Pfeiffer, J. A., Hart, J. L., Wood, L. A., Bhatnagar, A., Keith, R. J., Yeager, R. A., et al. (2021). The importance of urban planning: Views of greenness and open space is reversely associated with self-reported views and depressive symptoms. *Population medicine*, *3*, 20. https://doi.org/10.18332/popmed/139173