

Journal Pre-proof



ChatGPT in Scientific Writing: A Cautionary Tale

Haoyi Zheng M.D., Ph.D. , Huichun Zhan M.D.

PII: S0002-9343(23)00159-6
DOI: <https://doi.org/10.1016/j.amjmed.2023.02.011>
Reference: AJM 17081

To appear in: *The American Journal of Medicine*

Received date: 14 February 2023
Accepted date: 14 February 2023

Please cite this article as: Haoyi Zheng M.D., Ph.D. , Huichun Zhan M.D. , ChatGPT in Scientific Writing: A Cautionary Tale, *The American Journal of Medicine* (2023), doi: <https://doi.org/10.1016/j.amjmed.2023.02.011>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier Inc.

ChatGPT in Scientific Writing: A Cautionary Tale

Haoyi Zheng¹, M.D., Ph.D.; Huichun Zhan^{2,3}, M.D.

¹Saint Francis Hospital, The Heart Center, Roslyn, NY

²Department of Medicine, Stony Brook School of Medicine, Stony Brook, NY

³Medical Service, Northport VA Medical Center, Northport, NY

Correspondence:

Haoyi Zheng, MD, PhD, FACC

St Francis Hospital, The Heart Center

100 Port Washington Blvd,

Roslyn, NY 11576

Email, Haoyi.Zheng@chsli.org

Total word count: 1396 (including references)

Declarations of interest: none

Funding: None

Both authors had access to the data and participated in the preparation of the manuscript.

ChatGPT (OpenAI, San Francisco CA), an artificial intelligence chatbot, has garnered worldwide excitement since its release in November 2022.¹ Despite its growing popularity, concerns have also been raised about its impact on scientific writing and publishing. ChatGPT automatically creates text based on written prompts from users and is capable to generate various forms of writing, including essays, poems, lyrics for the song, and even an academic essay. Reports have emerged of articles in journals that were written by using ChatGPT,^{2,3} including a

recent article in the highly-regarded journal *Radiology*, which was entirely written by ChatGPT with a radiologist in training listed as the corresponding author.³ The article's headings and subheadings were used as prompts by the actual author to generate all the contents for the article. The increasing use of ChatGPT in scientific writing poses an unprecedented and immediate challenge to scientific publishing, as there are currently no guidelines or regulations in place to guide or govern its use. Major concerns are focused on copyright, attribution, plagiarism, and authorship for the articles generated by ChatGPT. There has been limited discussion to date regarding the accuracy of the content generated by ChatGPT.⁴

To evaluate the output of writing generated by ChatGPT, we used a recent article titled "it is time to abandon the use of body surface area (BSA) derived from a 100-year-old formula", which was written by our first author and published in 2022.⁵ This article should not be in ChatGPT's database, as its training data only goes up until 2021.¹ We first summarize key facts from the article and then repeatedly prompted ChatGPT with a question to assess its response.

The following are basic facts from the article:

1. In 1916, Du Bois Du Bois introduced the first BSA formula. $BSA(m^2) = \text{Weight}(kg)^{0.425} \times \text{height}(cm)^{0.725} \times 0.007184$. The article was published in the *Archives of Internal Medicine*.⁴ The formula was derived from 9 individuals.⁶ The formula has been used to calculate the BSA based on weight and height.

2. In 1987, Dr. Mosteller introduced a simplified formula for BSA calculation.⁷ $BSA (m^2) = ([Height(cm) \times Weight(kg)] / 3600)^{1/2}$. The article was published in the *New England Journal of Medicine*. The formula has been widely used due to its simplicity.

The facts were re-examined by prompting an open-ended question to ChatGPT: “How was the Du Bois formula generated?” The same question was prompted 5 times to regenerate responses. The output of ChatGPT was evaluated. Excerpts from each of the five responses along with our comments are presented below in order. For the full texts and screenshots from ChatGPT, please refer to the supplemental file.

1. Excerpt one: “Du Bois collected height and weight data from over 600 individuals, which he used to develop a mathematical equation to estimate BSA”
Comment: This is a clear falsification. The Du Bois formula was actually derived from data collected from only 9 individuals.
2. Excerpt two: “The Du Bois formula uses the following equation to calculate BSA: $BSA = 0.20247 \times height^{0.725} \times weight^{0.425}$ ”
Comment: The formula is incorrect. In fact, only one of three responses generated by ChatGPT related to formula was correct.
3. Excerpt three: “Du Bois was a sociologist and a pioneering figure in the field of African American studies. He was interested in understanding the social, economic, and political conditions that affected the health of African Americans and saw the importance of accurate measurements of body surface area in this regard.”

Comments: This is a fabricated tale-like statement generated by ChatGPT. Moreover, Du Bois's profession varied among physician, researcher, and sociologist in the different responses.

4. Excerpt four: "The Du Bois formula was generated by collecting height and weight data from a sample of healthy individuals and using statistical methods to determine the relationship between the two variables."

Comment: Statistical methods were not used by Du Bois at that time.⁶

5. Excerpt five: "The formula was created as a result of Dr. Du Bois's interest in improving the accuracy of BSA calculations. He believed that the commonly used Mosteller formula, which was based solely on height and weight, was inaccurate in certain populations, particularly those who were overweight or obese."

Comment: The Mosteller formula was introduced in 1987, 71 years after Du Bois formula.

All the five responses produced by ChatGPT are well-written and plausible-sounding. However, each response contains information that is either fundamentally wrong or fabricated. These falsifications (distorting data or findings) and fabrications (inventing data or cases) are not easily discernable to readers or inexperienced reviewers. ChatGPT simply extracts relevant data through literature searches, processes them, then creates its own story without considering the logic or accuracy of the story. This clearly indicates that the current version of ChatGPT is not ready to be used as a trusted source of information for scientific writing. Scientific writers who rely on ChatGPT must manually check all the "facts", statements, and references generated by ChatGPT. Therefore, there is no obvious advantage to writing with ChatGPT.

Scientific research is a human endeavor to uncover the truth and must be conducted in accordance with the highest ethical standards. The publication with incorrect data diverts the research from the truth and spreads misinformation. Therefore, careful data management is crucial as it directly reflects the integrity of researchers. Poor data management is scientific misconduct and should be avoided at all costs. Fabricating or falsifying findings is a serious scientific fraud and a violation of ethical standards in scientific research. Such misconducts can lead to sanctions against researchers and can even result in the end of their academic careers. For a ChatGPT-generated article, all listed authors are directly responsible for the accuracy and integrity for the output and are accountable for any misconduct. ChatGPT cannot be held accountable for any misconduct.

As demonstrated by our example, identifying fabrication or falsification during the peer-review process of a manuscript containing text generated by ChatGPT will pose a significant challenge for reviewers and editors. This challenge becomes even more worrisome if the actual authors of paper are not responsible for careful fact-checking. A recent study revealed that human reviewers could miss up to 32% of abstracts that were fully fabricated by ChatGPT during a consciously screening process.⁸ Scientific journals that accept ChatGPT's involvement in writing will face a significant increase in the number of retractions of published articles and a loss of credibility for the journals. Moreover, directly adopting the full text written by ChatGPT may constitute plagiarism and violate the code of conduct for scientific publishing, as originality is the foundation of scientific writing. Work that relies solely on ChatGPT's outputs lacks critical

thinking and reasoning skills of a human being, and can potentially be detrimental to the research and impede the science advancement.

In conclusion, while ChatGPT may hold endless possibilities for the future, its current form is far from mature to handle scientific writing. The future role of a more advanced ChatGPT in scientific writing will require comprehensive discussions and debates. The Science family of journals recently banned all ChatGPT-generated text, figures, images, or graphics. A violation of these policies will constitute scientific misconduct no different from altered images or plagiarism of existing works.⁹ Given the potential implications of fabricated and inaccurate information generated by ChatGPT, such a policy should become standard practice for all scientific publishing. Nonetheless, ChatGPT can still be a useful tool for checking grammar and syntax errors and refining the language, particularly for non-native speakers.

References

1. Open AI ChatGPT 2022. Accessed February 12, 2023, <https://openai.com/blog/chatgpt/>.
2. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 2023;613:620-1.
3. Biswas S. ChatGPT and the Future of Medical Writing. *Radiology*. Published online: February 6, 2023.
4. Liebrez M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *The Lancet Digital Health* 2023.
5. Zheng H. It Is Time to Abandon the Use of Body Surface Area Derived From a 100-Year-Old Formula. *Am J Med* 2022;135:e308-e10.
6. Shuter B, Aslani A. Body surface area: Du Bois and Du Bois revisited. *Eur J Appl Physiol* 2000;82:250-4.

7. Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med* 1987;317:1098.
8. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv* 2022:2022.12. 23.521610.
9. Thorp HH. ChatGPT is fun, but not an author. *Science* 2023;379:313-.

Journal Pre-proof