Peer Review File

Reviewer A

Comment 1: In my opinion, this paper is incomplete making the reproducibility less than desired. The authors functionally ran the product, report the output (lines 102 to lines 321), and then provide discussion. The authors should revise their approach to be reliably reproducible.

ChatGPT output can be Appendix at the end. Provide the original text and show with markup how the authors would alter.

Reply 1: In this brief experiment, we have highlighted the advantages and disadvantages of using a chatbot as a potential tool in medical writing. Consequently, the approach employed has several limitations in terms of reproducibility. This aspect has been extensively discussed in a dedicated paragraph (paragraph 4.1). The text produced by ChatGPT is presented in italics within the results section (paragraph 3.1), allowing for clear identification. In order to enhance the comprehension of the article's purpose, we have included three appendices: Appendix 1 provides the prompts utilized, Appendix 2 presents the checklist employed for text evaluation, and Appendix 3 showcases examples of erroneous outputs concerning references. Additionally, in this revised version, we have expanded upon the explanation of the methods employed, offering a more comprehensive overview. Finally, we have provided the criteria used for reproducibility (see reply to comment 3).

Changes in the text: Changes have been made throughout the entire text.

Comment 2: In Table 1, the authors provide notes, but not the criteria followed for those notes. In other words, a qualitative analysis has been performed, but methods for rigorous qualitative analyses were not followed. Different authors would produce different Table 1. Because the authors are analyzing a text, use a checklist and associated methods for qualitative study. Assure that with the text and the described Methods, another set of investigators would produce the same result.

Reply 2: In order to enhance the reproducibility of our study, we have taken measures to provide additional information and documentation. One of these measures is the inclusion of an appendix dedicated to the prompts and outputs used in the experiment (Appendix 1). In this appendix, readers will find a

comprehensive list of the prompts that were presented to the chatbot, along with the corresponding outputs generated by the chatbot in response to each prompt. Furthermore, we have also outlined the criteria used for evaluating the quality and accuracy of the text generated by the chatbot. These evaluation criteria, detailed in Appendix 2, can serve as a guideline for assessing the performance of the chatbot's responses.

Changes in the text: Prompts and outputs are reported in Appendix 1, checklist for text evaluation is in Appendix 2.

Comment 3: Table 2 provides percentages changed. However, the criteria used to create Table 2 are not given such that even if the authors had the identical text, they could produce a different product (lines 102 to 321). The criteria applied are needed for reproducibility.

Reply 3: We would like to thank the reviewer for this valuable suggestion. We have included the criteria used for the table in the text.

Changes in the text: The criterion used pertains to the word count in relation to the text produced by the prompts. These criteria include relevance, completeness, consistency and logical coherence (logical progression and cohesive structure), verification of information accuracy (reliable information), absence of redundancy (unnecessary repetition of information or ideas within the text), and clarity and readability.

Comment 4: The authors' Table 3 is a summary that could be obtained from multiple articles. These are without references.

Reply 4: According to this suggestion, references were added to the table.
Changes in the text: see Table 3.

Reviewer B
Thank you for your interesting addition to the scholarship on ChatGPT. It is an interesting paper and very well presents the promise and limitations of this emerging technology.

Some minor suggestions:
Abstract
Background

Comment 1: Line 36 - make "Artificial Intelligence" lower case as "artificial intelligence"
Reply 1: We corrected the mistake.

Comment 2: Line 37 - Delete "incredible" and consider changing to "It is a large language model capable of performing a variety..."
Reply 2: We corrected it as suggested.

Results
Comment 3: Line 47 - Change so it reads "Many important inaccuracies were found."
Reply 3: We corrected it

Conclusions
Comment 4: Line 48 - Consider changing first use of "important" to "significant."
Reply 4: we corrected it

Comment 5: Keywords: Lower case "intelligence"
Reply 5: we corrected it

Introduction
Comment 6: Great first paragraph. I recommend lower case for the abbreviated items unless they begin a sentence: Large language models (LLMs), artificial intelligence (AI), natural language processing (NLP) and so on throughout the text.
Reply 6: Thank you for your words. We corrected the mistakes.

Comment 7: Second Paragraph
Line 66
Suggest changing to "A widespread type of LLM is Chat Generative Pre-trained Transformer (ChatGPT), developed..."
Note that it is "pre-trained" not "pre-training."
Reply 7: Thank you, we revised the text

Comment 8: Also, I wasn't clear whether your intention was to leave untouched all that ChatGPT produced (including grammatical/punctuation errors, e.g., in Productive Models for Outcomes, Line 278 - "AI can is also employed..." - should be "AI is also employed..."; there is a stray close bracket in Education and Training, Line 306) to illustrate inadequacies large and small.
There is inconsistent punctuation (capitalization and punctuation - comma and

period placement before and after quotation marks) throughout the paper and tables.

Reply 8: As we intended to conduct text editing, we have performed a more comprehensive grammatical revision in this new version. Nevertheless, we have made sure to highlight and acknowledge this limitation in text production within the document.

Changes in the text: The output was integrated, edited (including revision of grammatical/punctuation errors), ...

Comment 9: Finally, in my limited experience, ChatGPT hasn't provided references, which is an obvious disadvantage in scientific writing. It was interesting to read that the model provided references but they were unreliable.

Reply 9: we inserted examples in Appendix 3.

Changes in the text: In Table 2: * The DOI precisely corresponds solely to the journal title (See Appendix 3).

Reviewer C

The authors reported an experimental paper on the use of chatGPT- a natural language program that enables the writing of scientific paper. This is an important topic.

Comment: My major concern was that I was not able to understand what type of experiment took place. My guess is that the authors used the chatGPT to write a part of the manuscript, but this was not very clear.

Another concern was that the evaluation process for the product of chatGPT was very brief and I am not sure if there was much of any evaluation.

Furthermore, the results section did not seem to contain any qualitative or quantitative evaluation of how the product prepared by chat GPT looked like for the authors. Was there any typos, grammatical errors, or inaccurate or omitted citations? Did the writing seem scientifically sound? Was it recognizable that the AI created this document?

Very interesting study!

Reply: Thank you for the words of appreciation. The purpose of the article was to assess ChatGPT's writing capabilities in a topic of particular relevance: AI in anesthesia. The text generated by ChatGPT is reported in italics in the results

section (paragraph 3.1). To better convey the article's objective to readers, we have added three appendices: Appendix 1 includes the prompts used, Appendix 2 contains the checklist used for text evaluation, and Appendix 3 presents examples of incorrect outputs related to references. Furthermore, in this new version, we have provided a more detailed explanation of the methods used.