# Qeios

Peer Review

# Review of: "When Algorithms Decide: The Controversial Role of AI in Cancer Care Auditing"

**Gustavo A. Patino**[1]

1. Homer Stryker MD School of Medicine, Western Michigan University, United States

In the manuscript by Yoshinari Júnior and Vitorino, the authors raise important points about the role of digital systems using artificial intelligence (AI) methods (in this case, large-language models specifically) in auditing oncology management plans. They illustrate their points by reviewing the responses provided by two versions of the LLM ChatGPT to a couple of clinical vignettes and close by . The topic is timely and important, as medicine (like so many other fields) is figuring out the proper role of these AI systems in supporting the decisions made by physicians. I have the following suggestions to improve the clarity and applicability of the paper:

- The main limitation of the paper is the use of only two vignettes. Please consider exploring the issues you present more widely with the use of additional cases. Do you get the same results with other types of cancers or with cases that are between the two you present? It might help to have cases where all the information is the same and only a single data point is changed.

- To improve the reproducibility of the data, the authors should provide more detail about their experiments with ChatGPT. These should include:

  - The dates of the experiments

  - The exact prompts used and any iterative modifications to them, as the quality of prompts is recognized as a significant determinant of the system output.

  - How many times the two ChatGPT versions were tested on each vignette and what the consistency of the outputs was. Given that LLMs are non-deterministic, this is critical information for both the internal and external validity of the findings, especially for the issue of the system providing an "unethical denial" as described in the paper.

- If any of the settings in the system were modified, especially the temperature, which determines how much randomness is allowed in the answers.
- Include images of the representative outputs for each experiment

- Table 1, with recommendations for ethical and effective integration of AI, represents the part of the paper that readers will come back to multiple times. The recommendations presented are good, but because they are confined to the table, they feel underdeveloped. I would suggest making them the focus of the second half of the manuscript and discussing them in more detail. If the authors need more space for this, I would suggest making the cases a table or figure and the output of the "unethical denial" a figure.

- Please add the exact references for the evidence and clinical guidelines on which the recommendations from Table 1 are based.

- From what I understood, a key message of the piece is that AI-based systems should not replace humans as the final decision-makers for high-stakes decisions. However, I felt this point sometimes got confused with the other important issue of access to the latest and validated AI systems, particularly in the Conclusion section when the authors talk about fairness and reliability questions due to inequities in access. I think in the Conclusion section, it should be made explicit that both, AI systems should not completely replace human decision-making and that inequities in access are a second issue to be addressed, like the authors preview in the Introduction.

## Declarations

**Potential competing interests:** No potential competing interests to declare.