

Peer Review File

Article Information: <https://dx.doi.org/10.21037/-23-635>

Reviewer A

This is a well-written and insightful review of the study by Musheyev et al., which evaluated the performance of AI chatbots in providing information on urological malignancies. The author provides a clear and concise summary of the study's findings and highlights both its strengths and limitations. The recommendations for future research are well-reasoned and should be valuable to researchers and developers working in this area.

Reply: Thank you.

Changes in the text: None

Here are some specific comments:

The point about the dynamic nature of LLMs is well-taken. It is important to consider how these models can be evaluated in a way that captures their ability to interact with users and provide personalized responses.

Reply: Thank you.

Changes in the text: None

The suggestion to educate users about how to interact with AI chatbots is also important. This could help to improve the understandability and actionability of the information provided by these models.

Reply: Thank you.

Changes in the text: None

The call for more comprehensive approaches to understanding quality, actionability, and understandability is welcome. This is an area where further research is needed.

Reply: Thank you.

Changes in the text: None

The emphasis on the need for a cautious yet inquisitive stance towards AI integration is important. It is important to carefully consider the potential risks and benefits of this technology before it is widely adopted in healthcare.

Reply: Thank you.

Changes in the text: None

The concluding point about the collaborative effort needed to shape the future of AI is well-stated. This is a complex issue that will require input from a variety of stakeholders.

Reply: Thank you.

Changes in the text: None

The author could consider adding a brief discussion of the potential impact of AI chatbots on the patient-doctor relationship.

Reply: Thank you for taking the time to read and provide feedback on our editorial, we agree and believe it could benefit from an extended discussion of ethics and impact of AI on the doctor-patient relationship.

Changes in the text: We have expanded our discussion of ethics (final paragraph before Conclusion).

The author could also explore the ethical considerations of using AI chatbots in healthcare, such as the issue of bias and transparency.

Reply: We agree and we believe the editorial could benefit from an extended discussion of ethics including bias and transparency.

Changes in the text: We have expanded our discussion of ethics (final paragraph before Conclusion).

Reviewer B

The article discusses the emerging role of large language models like ChatGPT in everyday life, particularly in the field of urology, raising concerns about accuracy and accountability. The editorial focuses on reporting mainly 2 studies describing urological queries.

Reply: Thank you.

Changes in the text: None

Good sides:

- The article is well written, coherent and linguistically correct
- The research analysis seems accurate
- Good summary

Reply: Thank you.

Changes in the text: None

Problems:

- The title and introduction should clearly specify what target group the article is for - a urologist / health professional or a patient.

Reply: We agree and have made this change in the manuscript

Changes in the text: Title was changed and phrase added to 3rd sentence of introduction paragraph 2.

- The introduction explains that we evaluate questions for ChatGPT, but it is not specified which ones - whether it is about symptoms, patient's questions, or urologist's doubts.

Reply: We agree further information in the introduction regarding the queries strengthens the editorial. The exact queries are found in Supplementary Table 1 of Musheyev et al., and contain a range of queries including symptoms, questions, definitions, and singular words.

Changes in the text: A sentence was added to the introduction clarifying where the queries came

from (second sentence of introduction paragraph 2).

- It should be indicated on which population / group / number of questions the research was performed

Reply: We agree, clarification regarding exact groups/questions/methods in Musheyev et al., is necessary.

Changes in the text: Clarification added to first paragraph of Study Summary section.

- The results are difficult to understand because most urologists do not know the scales, used in articles, an explanation, e.g., in the form of a table, would be useful. Additionally, cut-off thresholds, if any, should be described (e.g., the tool is considered reliable for a result of at least 80%).

Reply: We agree and have added this table to our manuscript including ranges and interpretation, however cutoff thresholds were not applicable or available for most of these metrics included.

Changes in the text: We added a table with the scales used in the article and reference the table in the Study Summary first paragraph.