Peer Review File

Article Information: https://dx.doi.org/10.21037/tau-23-556

Reviewer A

This is a brief Editorial Commentary about AI-driven health communication. I have some suggestions for improvement:

"Medical informatics has witnessed increasing interest, largely directed towards the capabilities of artificial intelligence (AI) and its potential to redefine how healthcare professionals communicate with their patients." -> I think the influence of AI on medical informatics is much larger than professional-patient communication; this is only one of many applications of AI in this field.

Reply: Thank you for this important point.

Changes in text: We deleted the that sentence and replaced with "Medical informatics has experienced a significant surge in interest, especially regarding the diverse capabilities of artificial intelligence (AI). These advancements are not limited to transforming physician-patient communication but extend to various other aspects of healthcare, including diagnostics, treatment planning, and healthcare management."

"AI-driven Large Language Models (LLMs)" -> I think LLMs drive AI, not the other way around.

Reply: Thank you. Yes, I think the distinction hinges on how one interprets AI-driven. Whether as a broad field that includes various tech such as machine learning, or if we see LLMs as an integral dricing force in the field of AI. The title of the paper we are commenting on uses AI-powered LLMs.

Changes in text: We made changes to avoid using this term. and will stick to using either AI or LLMs.

- Please include a brief explanation on what LLMs are.

Reply: Thank you.

Changes in text: "Large Language Models (LLMs) are designed to understand and generate human language. They learn from a vast amount of text data, enabling them to answer questions, write texts, and perform various language-related tasks. LLMs such as ChatGPT and BERT have gained attention as potential alternatives to widely used search engines like Google. Their potential use for medical queries highlights the need to evaluate their accuracy and reliability in giving medical advice."

- line 90: "typing writing"?
Reply: Thank you.
Changes in text: Deleted "writing"

Some issues are not mentioned:
- reliability and accountability of results: what if the LLM produces nonsense?
- many patients still prefer to talk to humans instead of an AI
Reply: Definitely. We added these crucial points.
Changes in text: "In addressing the reliability and accountability of LLM-generated results, it's crucial to consider scenarios where the model may produce nonsensical or inaccurate information. Despite their advanced algorithms, LLMs are not infallible and can generate errors or misleading content (10). This highlights the need for mechanisms to identify and correct such instances promptly. Lastly, the preference of many patients to interact with human healthcare providers rather than AI systems must be acknowledged. The human element in medical care, encompassing empathy, understanding, and personal judgment, remains irreplaceable. As of now, while LLMs can be valuable tools, they should complement, not replace, human medical expertise, ensuring that patient care remains grounded in personal interaction and professional judgment."

Reviewer B

Generally, a good and concise paper. I would improve it only by making it clear which LLM(s) you are referring to, otherwise, make it clear that you are referring to all LLMs, or a specific range of them (e.g., ChatGPT, BERT & roBERTa).
Reply: Thank you. Yes, we agree.
Changes in text: We elaborated on explaining what LLMs are and then specified that we are referring to LLMs broadly.

I don't think that the possibility of misinformation was touched upon in enough depth. LLMs can hallucinate data, so this should be included as a risk. One way of avoiding this would be to ensure the LLM is trained on specific, medically accurate data, but the above mentioned LLMs are not specifically trained on just these. Just something to mention in a bit more depth.
Reply: Absolutely. We have included morein the text.
Changes in text: In addressing the reliability and accountability of LLM-generated results, it's crucial to consider scenarios where the model may produce nonsensical or inaccurate information. Despite their advanced algorithms, LLMs are not infallible and can generate errors or misleading content (10). This highlights the need for mechanisms to identify and correct

such instances promptly.

Reviewer C

The authors should be commended on this summary report
Reply: Thank you very much.
Changes in text: None.

Reviewer D

The manuscript comments on the study “Evaluating the Effectiveness of Artificial Intelligence-powered Large Language Models Application in Disseminating Appropriate and Readable Health Information in Urology” published at Journal of Urology. The authors emphasize the necessity for better communication using language accessible to the major public and how AI can make urology health care easier to digest. The paper shares the author's optimistic opinion on a currently relevant topic in medicine in general. The text is well-written and meets the objective of this type of scientific production. Minor comment – in line 90, eliminate either the word “typing”or the word “writing.” Otherwise, it is easy to read, understand, and process.
Reply: Thank you very much.
Changes in text: We eliminated the word "writing"

Reviewer E

This is a well-written editorial commentary on a topical issue. Authors highlight some key areas under discussion in this field, though I would have liked to see a more extensive review of current literature and thought leadership in the domain of LLM evaluation. Was an LLM utilized in the drafting or writing of this commentary, and if so, was this declared?
Reply: Thank you. We used ChatGPT-4 to review our writing for grammar and organization. We added the declaration, thought leadership and a provided another review of a paper (10).
Changes in text: In addressing the reliability and accountability of LLM-generated results, it's crucial to consider scenarios where the model may produce nonsensical or inaccurate information. Despite their advanced algorithms, LLMs are not infallible and will generate misleading content and hallucinate data (10). This highlights the need for mechanisms to identify and correct such instances promptly. Lastly, the preference of many patients to interact with human healthcare providers rather than AI systems must be acknowledged. The human

element in medical care, encompassing empathy, understanding, and personal judgment, remains irreplaceable. As of now, while LLMs can be valuable tools, they should complement, not replace, human medical expertise, ensuring that patient care remains grounded in personal interaction and professional judgment.

Acknowledgement: ChatGPT-4 was utilized to review our writing for grammar and organization.

However, in the pursuit of simplicity, there is a risk of sacrificing comprehensive information for clarity. This underscores the importance of balancing the need for understandable explanations with the necessity of conveying accurate and complete information. Achieving this balance is a task that requires the art side of medicine, which involves tailoring information to the unique needs and understanding of individual patients. In a survey study of patients with prostate cancer, patients stated that they had higher trust in a diagnosis made by AI controlled by a physician versus AI not controlled by a physician. Furthermore, AI-assisted physicians were preferred over physicians alone (8).

Reviewer F

Page 3, line 90: "by typing writing" simply sounds odd. Please rephrase to one of these verbs.
Reply: Thank you for pointing out this error.
Changes in text: We deleted "writing"

Page 3, line 73-77: I would combine this paragraph with the previous one given the length.
Reply: Thank you.
Changes in text: We combined the two paragraphs.

Page 3, line 90-92: It is important to note whether there is any sacrifice in accuracy of information for the purpose's clarity of presentation thereof. Please address this.
Reply: Thank you. This is an important point.
Changes in text: However, in the pursuit of simplicity, there is a risk of sacrificing comprehensive information for clarity. This underscores the importance of balancing the need for understandable explanations with the necessity of conveying accurate and complete information. Achieving this balance is a task that requires the art side of medicine, which involves tailoring information to the unique needs and understanding of individual patients.

Page 5, lines 115-119: The subheading of "Ensuring Accuracy" with the initial two paragraphs align thematically with this subheading. The ethical repercussions of patient health privacy,

while important, is perhaps miscategorized as it does not align thematically. It may be better suited for the previous section or ought to be ameliorated into the body in another fashion. It does not appear that this is a concern in your concluding paragraphs as well despite being a strong point.

Reply: Thank you.

Changes in text: We changed the subheading to "Ensuring Accuracy and Safety". We also elaborated more on the last paragraph: "In addressing the reliability and accountability of LLM-generated results, it's crucial to consider scenarios where the model may produce nonsensical or inaccurate information. Despite their advanced algorithms, LLMs are not infallible and will generate misleading content and hallucinate data (11). This highlights the need for mechanisms to identify and correct such instances promptly. Lastly, the preference of many patients to interact with human healthcare providers rather than AI systems must be acknowledged. The human element in medical care, encompassing empathy, understanding, and personal judgment, remains irreplaceable. As of now, while LLMs can be valuable tools, they should complement, not replace, human medical expertise, ensuring that patient care remains grounded in personal interaction and professional judgment."

Reviewer G

Well written editorial commentary regarding an important recent J Urol paper. Highlights important points. AI has positive and negative comments and this commentary highlights the need for regulation and monitoring to ensure it is utilized appropriately

Reply: Thank you.

Changes in text: None.

Reviewer H

This is article is an editorial on the manuscript:

“Evaluating the Effectiveness of artificial intelligence-powered large language model application in disseminating appropriate and readable health information in urology”

The authors do a nice job highlighting the “promise” of AI and the potential benefits that can be offered to patients within the context of the manuscript.

There is a somewhat limited discussion on the limitations of AI.

Reply: Thank you. We included writing on various limitations of AI throughout the text.

Changes in text: However, in the pursuit of simplicity, there is a risk of sacrificing comprehensive information for clarity. This underscores the importance of balancing the need for understandable explanations with the necessity of conveying accurate and complete

information. Achieving this balance is a task that requires the art side of medicine, which involves tailoring information to the unique needs and understanding of individual patients. In a survey study of patients with prostate cancer, patients stated that they had higher trust in a diagnosis made by AI controlled by a physician versus AI not controlled by a physician. Furthermore, AI-assisted physicians were preferred over physicians alone (8).

In addressing the reliability and accountability of LLM-generated results, it's crucial to consider scenarios where the model may produce nonsensical or inaccurate information. Despite their advanced algorithms, LLMs are not infallible and will generate misleading content and hallucinate data (11). This highlights the need for mechanisms to identify and correct such instances promptly. Lastly, the preference of many patients to interact with human healthcare providers rather than AI systems must be acknowledged. The human element in medical care, encompassing empathy, understanding, and personal judgment, remains irreplaceable. As of now, while LLMs can be valuable tools, they should complement, not replace, human medical expertise, ensuring that patient care remains grounded in personal interaction and professional judgment.

Recommendations:

1) Given that patients would be using these tools to receive what they assume to be accurate information, it must be more explicitly stated that these tools are not yet ready to be implemented as the accuracy of these tools is nowhere near an acceptable level for prime time.

2) Line 90 - grammatical error: omit the word writing

Reply: Thank you. We omitted the word writing and also added the following:

Changes in text: In addressing the reliability and accountability of LLM-generated results, it's crucial to consider scenarios where the model may produce nonsensical or inaccurate information. Despite their advanced algorithms, LLMs are not infallible and will generate misleading content and hallucinate data (11). This highlights the need for mechanisms to identify and correct such instances promptly. Lastly, the preference of many patients to interact with human healthcare providers rather than AI systems must be acknowledged. The human element in medical care, encompassing empathy, understanding, and personal judgment, remains irreplaceable. As of now, while LLMs can be valuable tools, they should complement, not replace, human medical expertise, ensuring that patient care remains grounded in personal interaction and professional judgment.