Peer Review File

Reviewer A

1. I suggest reviewing the first sentence from a grammatical point of view, for example, "Artificial intelligence has long captivated the imagination. Most recently, a number of large language models (LLMs) have been launched, including ChatGPT (OpenAI/Bing), Bard (Google/Alphabet), and LLaMA (Meta).
   Reply: this was adjusted
   Changes in text: changes as above

2. It could be interesting to compare the results of the authors' research with the ones of a similar article, "Quality of information and appropriateness of ChatGPT outputs for urology patients," DOI: 10.1038/s41391-023-00754-3.
   Reply: This article was included in discussion
   Changes to text: In a similar project, 100 urology case studies were presented to ChatGPT 3.5 and found only 52% of responses appropriate. This suggests that the models have a long way to go for trustworthy direct patient facing clinical responses (3).

3. In general, to make the editorial more interesting, it would be nice to include information and opinions on the role of artificial intelligence obtained from the numerous articles available in literature.
   Reply: More articles were included. However this is meant to be a short commentary on a specific article and as such, expanding substantially felt incorrect. I did expand discussion of novel ways to improve chat bots in a urologic specific manner.
   Changes to text: Additionally, a process termed retrieval-augmented generation has LLMs use specific datasets or trusted technical or policy documents as external sources that can be cited by the LLM output. For instance, a urology specific chat bot was developed based on the EAU Oncology guidelines. When built directly on guidelines, it is no surprise that accuracy of response to patient queries improved (4).

Reviewer B

A reasonable commentary, but misses an opportunity for a more in-depth discussion on certain arenas, such as the ongoing discussion around evaluation of LLM output and the use of AI to generate training data for AI models. I have noted a number of papers/commentaries comment on the readability of LLM output, this of course as the authors of this editorial have mentioned, are easily amended through appropriate prompts/meta-prompts. It is a good opportunity to highlight such misconceptions and the stochastic nature of many of these models. The

emerging trend of RAG-models should also be discussed, instead of just fine-tuning, which is computationally and financially expensive.

Reply: excellent suggestion for the RAG model and citation. This has been done

Changes to text: Additionally, a process termed retrieval-augmented generation has LLMs use specific datasets or trusted technical or policy documents as external sources that can be cited by the LLM output. For instance, a urology specific chat bot was developed based on the EAU Oncology guidelines. When built directly on guidelines, it is no surprise that accuracy of response to patient queries improved (4).

It is a contemporary and well-written editorial that excellently introduces the various aspects of using artificial intelligence in medicine, particularly in urology. I have no objections and recommend the acceptance of the editorial.

Perhaps it could be added that the next step for LLMs (Large Language Models) doesn't solely lie in fine-tuning the models or improving algorithm training, but also in the development of domain-specific LLMs. In urology, there already exists such a model called 'Uro_Chat,' which has been published and validated (Khene ZE, Bigot P, Mathieu R, Rouprêt M, Bensalah K; French Committee of Urologic Oncology. Development of a Personalized Chat Model Based on the European Association of Urology Oncology Guidelines: Harnessing the Power of Generative Artificial Intelligence in Clinical Practice. Eur Urol Oncol. 2023 Jul 18:S2588-9311(23)00139-6. doi: 10.1016/j.euo.2023.06.009. Epub ahead of print. PMID: 37474402. // May M, Körner-Riffard K, Marszalek M, Eredics K. Would Uro_Chat, a Newly Developed Generative Artificial Intelligence Large Language Model, Have Successfully Passed the In-Service Assessment Questions of the European Board of Urology in 2022? Eur Urol Oncol. 2023 Sep 14:S2588-9311(23)00178-5. doi: 10.1016/j.euo.2023.08.013. Epub ahead of print. PMID: 37716835.).

Reply: I appreciate the references and did include the most relevant.

Changes to text: For instance, a urology specific chat bot was developed based on the EAU Oncology guidelines. When built directly on guidelines, it is no surprise that accuracy of response to patient queries improved (4).