Critical Care

COMMENT

Open Access

Beware the self-fulfilling prophecy: enhancing clinical decision-making with AI

Taotao Liu[1] and Yaocong Duan[2*]

The integration of artificial intelligence (AI) in healthcare through large language models (LLMs) presents a paradigm shift with both promising opportunities and significant challenges. As we stand at the cusp of this transformation, it is crucial to critically examine its implications, particularly the potential pitfalls. Here, we discuss the latent risk of over-reliance on AI-driven clinical decision-making and propose potential solutions to mitigate this risk. While acknowledging the potential of LLMs to enhance healthcare delivery, we emphasize the importance of implementing a post-decision review process and utilizing its outcomes to create curated, publicly available datasets. This approach aims to avoid a "self-fulfilling prophecy" where AI systems amplify existing biases. By establishing this feedback loop, we can synergize AI capabilities with the expertise of clinicians, ensuring responsible and effective integration of LLMs in healthcare.

Our recent study[1] evaluated the accuracy of GPT-4.0 in predicting the risk for endotracheal intubation within 48 h for 71 patients receiving high-flow nasal cannula (HFNC) oxygen therapy by comparing its performance to that of clinical physicians. The area under the receiver operating characteristic curve (AUROC) reflecting the predictive efficacy of GPT-4.0 has been comparable to that of specialist physicians (0.821 vs. 0.782, $p=0.475$) and superior to that of non-specialist physicians (0.821 vs. 0.662, $p=011$) [1]. Similar findings have been reported in other clinical scenarios [2–4], highlighting the potential of the near-future application of LLMs in augmenting clinical decisions and answering medical questions.

The integration of LLMs in clinical decision-making holds the promise of substantially increasing the efficiency of healthcare delivery and promoting the homogenization of healthcare services. By providing rapid, data-driven insights, LLMs can assist clinicians in making timely and accurate decisions, thus potentially improving patient outcomes [3, 5]. Moreover, AI can help standardize care practices across different healthcare settings and ensuring that patients receive consistent and high-quality care regardless of location and development levels.

However, these promising results also raise significant concerns. To illustrate, consider a hypothetical scenario from our study: a patient with respiratory failure receiving HFNC oxygen therapy has an AI-predicted 70% risk (95% CI: 60–85%) of requiring intubation within the next 48 h. Based on this prediction, delaying intubation could increase the patient's mortality risk, prompting physicians to opt for immediate intubation and mechanical ventilation. This decision would seemingly validate and reinforce the AI's prediction, leading to higher predicted risks for similar cases in the future since the AI's learning is based on real-world data outcomes. This situation resembles a self-fulfilling prophecy, akin to the narrative in Shakespeare's play "Macbeth." However, patients might have avoided unnecessary intubation had the clinical decision been based on comprehensive clinical judgment rather than AI prediction alone.

This comment refers to the article available online at https://doi.org/10.1016/j.heliyon.2024.e31750.

*Correspondence:
Yaocong Duan
yaocong.duan@glasgow.ac.uk
[1] Department of Surgical Intensive Care Unit, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing 100730, China
[2] School of Psychology and Neuroscience, University of Glasgow, Glasgow G12 8QQ, UK

Potential over-reliance on AI could lead to a loss of clinical intuition, reducing the clinician's role to merely following algorithmic recommendations without critical judgment. A potential solution to mitigate these concerns involves a post-decision review process where, after clinical decisions are made with AI involvement, all data are archived and reviewed by multiple specialist physicians. This review process can help correct any suboptimal judgments and create a refined dataset. Making this data publicly available can further enhance the reliability of AI systems. LLMs like GPT could then learn from these high-quality, curated datasets, potentially improving their predictive capabilities and making more accurate and reliable future predictions.

This solution leverages the collective expertise of specialist physicians to refine AI-driven clinical decisions, ensuring that AI recommendations are critically evaluated and validated. However, there are challenges, such as ensuring the privacy and security of patient data, obtaining consistent and thorough reviews from specialists, and addressing legal and ethical concerns for applications of AI in healthcare. For example, the European Union's General Data Protection Regulation (GDPR) requires that algorithms must have transparency before being used in patient care [5]. Despite these challenges, this approach presents a pathway towards a future where AI and clinicians work synergistically to enhance patient outcomes, improve healthcare delivery efficiency, and reduce disparities in healthcare services across different locations and development levels.

Acknowledgements
None.

Author contributions
Writing – original draft, T.L. and Y.D.; writing – review & editing, T.L. and Y.D.

Funding
T.L. was supported by National High Level Hospital Clinical Research Funding [BJ-2023–173].

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Competing interest
The authors declare no competing interests.

Received: 6 August 2024 Accepted: 7 August 2024
Published online: 26 August 2024

References

1. Liu T, Duan Y, Li Y, Hu Y, Su L, Zhang A. ChatGPT achieves comparable accuracy to specialist physicians in predicting the efficacy of high-flow oxygen therapy. Heliyon. 2024;10: e31750.
2. Cabral S, Restrepo D, Kanjee Z, Wilson P, Crowe B, Abdulnour R-E, et al. Clinical reasoning of a generative artificial intelligence model compared with physicians. JAMA Intern Med. 2024;184:581.
3. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620:172–80.
4. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on Medical Challenge Problems [Internet]. arXiv; 2023 [cited 2024 Jul 21]. Available from: http://arxiv.org/abs/2303.13375
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56.

Publisher's Note