

Large Language Model in Critical Care Medicine: Opportunities and Challenges

Sameera Hajjama¹, Deven Juneja², Prashant Nasa³

Keywords: Artificial, Artificial intelligence, Artificial neural network, Critical care medicine, Healthcare, Intelligence.

Indian Journal of Critical Care Medicine (2024); 10.5005/jp-journals-10071-24743

As artificial intelligence (AI) models continue to evolve and integrate into the healthcare system, it becomes vital to understand these innovative tools to maximize their potential and anticipate the risks. A large language model (LLM) is one of the largest forms of artificial neural networks that utilize algorithmic models to process and generate a natural language text resembling the human mind from user-generated prompts. Many are familiar with the famous model ChatGPT, but models such as Google Gemini are frequently utilized, and such models have been known to demonstrate creativity and precision.¹ The tools are popular due to human-like-output, widespread public availability, and ease of usability.

The potential applications of LLM in healthcare are numerous, ranging from patient care to education and research. Patient care can be improved through integrating medical knowledge and interpersonal communication.² Large language models can improve communication between healthcare providers and patients through accurate real-time translation in many languages, and converting medical jargon into layman, easy-to-understand language. Further, accurate documentation and reports from unstructured notes, automated dictation, and prompt-driven chart review could help to reduce physical and cognitive load and enhance the opportunity for patient-physician interactions.^{2,3} Large language model could facilitate research through a rapid and accurate summary of scientific manuscripts, paraphrasing scientific contents, democratizing research by overcoming language barriers, generating scientific abstracts, and interpreting data through coding.⁴

Finally, LLM has revolutionized education by providing summaries of scientific content, preparing effective presentations and engaging interactive simulations, step-by-step explanations of the concepts, and contextualization and translation of various topics in medicine in a personalized depth, tone, and style.^{3,4} Students use ChatGPT due to its ability to answer complex scenarios in simple terms and use questions to generate output to a student's level of understanding, allowing students to learn at their own pace. The tool has been proven to even pass the USMLE exam due to its ability to respond to intricate medical scenarios.^{2,5}

APPLICATION OF LLM IN CRITICAL CARE MEDICINE

Researchers and physicians are exploring the potential of LLM in the field of critical care medicine (CCM). A known benefit of LLM tools, including algorithmic models, is in the administrative tasks, documentation, and production of referral notes and discharge

¹Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

²Institute of Critical Care Medicine, Max Super Speciality Hospital, New Delhi, India

³Department of Critical Care Unit, The Royal Wolverhampton Trust, Wolverhampton, United Kingdom

Corresponding Author: Prashant Nasa, Department of Critical Care Unit, The Royal Wolverhampton Trust, Wolverhampton, United Kingdom, Phone: +447852862083, e-mail: dr.prashantnasa@hotmail.com

How to cite this article: Hajjama S, Juneja D, Nasa P. Large Language Model in Critical Care Medicine: Opportunities and Challenges. *Indian J Crit Care Med* 2024;28(6):523–525.

Source of support: Nil

Conflict of interest: None

summaries, thereby reducing extra hours of work required of a physician.⁴ Chat-GPT has been used in ICU to integrate patient information with data and provide accurate diagnosis for clinical-decision-making, early warning systems, diagnostic models from large databases and improve patient communication through layman text. Critical care practitioners can benefit from LLM's ability to aid in the diagnosis of medical conditions, provide a synthesis of relevant guidelines for evidence-based decision support, stratify the risk of patients undergoing high-risk procedures, such as extracorporeal membrane oxygenation (ECMO), and optimize assessment before such a procedure.⁶ They have also been used to organize unstructured clinical text and analyze large forms of data, extracting the relevant information.⁶

Large language models are revolutionizing critical care by contributing to the management of intensive procedures. A study evaluated the use of ChatGPT3.5 and ChatGPT4 for addressing alarms during continuous renal replacement therapy (CRRT). Both LLMs showed comparable accuracy and consistency in alarm troubleshooting. However, an improvement in reliability was recommended in future models. This has proven the ability of LLM tools to generate recommendations for CRRT management while avoiding outputs that may be harmful to patients.⁷

An important aspect of LLM tools is the ability to break down complicated topics into manageable chunks and create a simplified template to understand complex scenarios. This can be immensely helpful in the critical care field, as patients undergoing end-of-life

care (EOLC) require intensive medical management, and discussions with family or friends may be suboptimal due to a barrier to medical literacy in the general population. Artificial intelligence tools can help by breaking down complex medical scenarios into smaller patient information leaflets to optimize the ability of the general population to make crucial decisions for their loved ones.

In the current issue, Gondode et al., compared the two popular LLMs, ChatGPT and Google Gemini, for patient education by generating information leaflets on EOLC. Although both conveyed a positive sentiment with high levels of accuracy and completeness, Google Gemini PILs showed superiority in terms of readability and actionability, and inferiority in terms of low accuracy scores.⁸

LIMITATIONS AND CHALLENGES OF LLM

Although LLM tools provide innovative methods of data utilization, it is imperative to understand that the fundamentals of AI tools rely on algorithmic systems, so the performance of the data output is highly contingent on the quality of the input data. If the input data is incomplete, biased, or filled with errors, then the output will be similar.^{1,4}

Misinterpretation or existing medical biases in the training dataset will be reproduced or enhanced as the output will always be dependent on user prompts. Biases produced by inequalities in data representation due to factors such as gender, race, socioeconomic status, or sexual orientation may be perpetuated with these LLMs.⁹ ChatGPT 3.5 is trained in the dataset up to September 2021, and hence, the output may lack the recent changes which may produce scientifically incorrect output. Hallucinations of LLMs are another challenge, defined by the generation of misinformation that seems plausible but is not based on the real or existing data and is instead produced by the model interpretation or extrapolation of the data. This misinformation may inadvertently cause harm if the healthcare professional is unaware of AI hallucination. Hence, awareness of AI hallucinations among healthcare professionals is essential and verification of output is recommended before its direct application.¹⁰

Large language models can facilitate patient communication by creating instructions for patient information, language translation, and creating medical summaries. However, a lack of true empathy with machine-driven LLMs may limit their use in certain emotionally charged situations.⁴

There is also the ethical implication of risking patient confidentiality by exposing vital information to the data input and the risk of data breach.⁴ As of now, there is a lack of formal accountability or consequences for ethical output breaches, so further regulatory boundaries must be established. There is also a consideration of a lack of transparency as to how specific inputs come to certain conclusions, and it takes a skilled individual to be able to trace down such data. Hence, the lack of true understanding of the mechanism of LLMs and the black-box nature of LLMs on the training dataset because of parent companies also add to ethical concerns. The usage of LLM in patient care requires upholding patient trust, privacy, and confidentiality. Informed consent with full disclosure of the AI tools used, and a comprehensive review of the harms and benefits of using each tool, must be obtained from patients.¹¹

Large language model algorithms are also built on learned associations of quantitative data, so using LLM tools in interpreting qualitative data may not generate the output required for adequate

usage.⁴ Besides, lack of access to real-time datasets with certain LLMs or access to only publicly available data, may limit their output from the latest scientific updates. Hence, the application of LLMs by researchers needs to mitigate misinformation, lack of comprehensiveness, and reproducibility.⁴

Many systems, such as human biology are unfamiliar with LLM tools. When prompted to make complex connections, the algorithms may be unable to generate adequate outputs.¹²

CONCLUSION

Large language models are rapidly changing the landscape of patient care, medical research, and education. The potential applications of LLMs in CCM include enhancing diagnosis precision, improving clinical decision-making, reducing healthcare professional's workload, and improving patient-physician communication. However, awareness about the challenges and limitations of LLMs is urgently needed to mitigate the risk, and intensivists must keep abreast of the rapidly changing field of AI.

ORCID

Sameera Hajjama  <https://orcid.org/0009-0001-1591-9031>

Deven Juneja  <https://orcid.org/0000-0002-8841-5678>

Prashant Nasa  <https://orcid.org/0000-0003-1948-4060>

REFERENCES

1. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. *Commun Med (Lond)* 2023;3(1):141. DOI: 10.1038/s43856-023-00370-1.
2. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2023;2(2):e0000198. DOI: 10.1371/journal.pdig.0000198.
3. Williams CYK, Zack T, Miao BY, Sushil M, Wang M, Kornblith AE, et al. Use of a large language model to assess clinical acuity of adults in the emergency department. *JAMA Netw Open* 2024;7(5):e248895. DOI: 10.1001/jamanetworkopen.2024.8895.
4. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med* 2023;29(8):1930–1940. DOI: 10.1038/s41591-023-02448-8.
5. Mihalache A, Huang RS, Popovic MM, Muni RH. ChatGPT-4: An assessment of an upgraded artificial intelligence chatbot in the United States Medical Licensing Examination. *Med Teach* 2024;46(3):366–372. DOI: 10.1080/0142159X.2023.2249588.
6. Lu Y, Wu H, Qi S, Cheng K. Artificial intelligence in intensive care medicine: Toward a ChatGPT/GPT-4 Way? *Ann Biomed Eng* 2023;51(9):1898–1903. DOI: 10.1007/s10439-023-03234-w.
7. Sheikh MS, Thongprayoon C, Qureshi F, Suppadungsuk S, Kashani KB, Miao J, et al. Personalized medicine transformed: ChatGPT's contribution to continuous renal replacement therapy alarm management in intensive care units. *J Pers Med* 2024;14(3):233. DOI: 10.3390/jpm14030233.
8. Gondode PG, Khanna P, Sharma P, Duggal S, Garg N. End-of-life care patient information leaflets—A comparative evaluation of artificial intelligence-generated content for readability, sentiment, accuracy, completeness, and suitability: ChatGPT vs Google Gemini. *Indian J Crit Care Med* 2024;28(6):561–568.
9. Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. *NPJ Digit Med* 2023;6(1):195. DOI: 10.1038/s41746-023-00939-z.
10. Hatem R, Simmons B, Thornton JE. A call to address AI "Hallucinations" and how healthcare professionals can mitigate their risks. *Cureus* 2023;15(9):e44720. DOI: 10.7759/cureus.44720.

11. Gupta S, Juneja D, Garg SK. Uses of Artificial Intelligence in Intensive Care Units. In: Dixit SB, Chaudhry D, Todi SK (Eds). *Critical Care Update 2021*. Delhi: Jaypee Brothers Medical Publishers (P) Ltd.; 2022. pp. 343–347. DOI: 10.5005/jp/books/18529_69.
12. Ullah E, Parwani A, Baig MM, Singh R. Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology – A recent scoping review. *Diagn Pathol* 2024;19(1):43. DOI: 10.1186/s13000-024-01464-7.