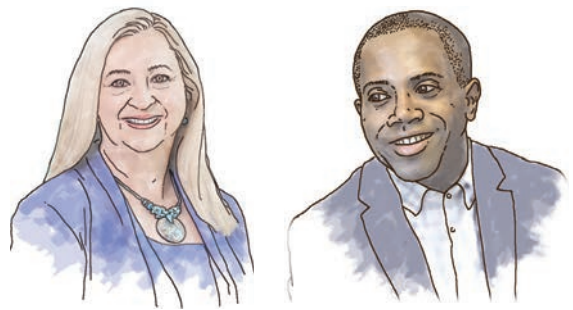


Editorial

ARTIFICIAL INTELLIGENCE IN CRITICAL CARE PRACTICE AND RESEARCH

By Cindy L. Munro, PhD, RN, ANP, and Aluko A. Hope, MD, MSCE



We believe that the purpose of clinical research is to understand clinical phenomena and to provide robust evidence for practice. In order to ensure the quality and integrity of the evidence upon which clinicians depend in caring for critically ill patients, every step of the research process must adhere to standards of accuracy, veracity, and ethical behavior. Researchers are responsible for the entire process—from the generation of the research question, through design and conduct of the study, analysis of results, to dissemination that facilitates translation to practice.

Critical care has a long history of incorporating technology into patient care. Contemporary critical care units look very different from those of 50 years ago. Nearly every aspect of care involves highly sophisticated equipment that can perform tasks delegated by direct care clinicians. For example, the pumps that precisely deliver intravenous fluids and alert nurses to problems are so ubiquitous that it is easy to forget that nurses originally adjusted intravenous flow rates manually, monitoring flow by counting drops per minute and using a time tape on the

bag of intravenous fluid (and alerted to problems by direct observation or a patient's complaint). The shift from manual to electronic charting not only led to more systematic documentation of patient care but also allowed inclusion of guidelines, prompts, and "nudges" to inform providers and influence providers' behavior.

Similarly, evolving technology has changed many research processes. As an example, reviewing the literature used to require in-person visits to a library to locate physical copies of journal articles in bound volumes of journals. The public release of PubMed in 1996 revolutionized access to primary and secondary sources underpinning research. PubMed is a database of biomedical literature created and maintained by the National Center for Biotechnology Information at the US National Library of Medicine, which is part of the National Institutes of Health.¹ Clinicians and researchers quickly realized the benefits of PubMed over physical searches at a library. The database was provided without any user fees, was accessible online, could be searched by multiple parameters (for example, keywords, author names, journal name, and date ranges), included journals that were judged to meet selected quality metrics, included abstracts and citation information, and provided links to outside sources for full-text

©2023 American Association of Critical-Care Nurses
doi:<https://doi.org/10.4037/ajcc2023958>

“ Chatbots cannot evaluate the quality or source of the data or of the response, and any bias or misinformation present in the data they use will be reflected in their responses. ”

articles. PubMed is an unqualified success story, now containing more than 35 million citations, and remains central to the work of clinicians and health care researchers. The availability of citation management software further assisted researchers in organizing and easily using the literature central to their work. The work of researchers has been further streamlined as word processing software replaced typewriters, data capture applications such as REDCap² replaced paper data collection forms, and statistical software replaced computer card stacks and hand calculations.

These technological improvements in clinical care and research all have in common that they were developed and programmed by humans, have operational rules that are understandable to their programmers (if not always to the end users), and yield predictable outputs. As such, these are examples of “white box” technologies. We have confidence in these tools in large part because we can “see” how they work and can usually detect when their performance is suspect.

In November 2022, ChatGPT³ burst onto the scene as a publicly available technology. ChatGPT is different from the technologies we describe above. ChatGPT is an artificial intelligence (AI) large language model (LLM) designed by the research lab OpenAI.³ Other LLM chatbots have been released following ChatGPT, including Microsoft’s Bing and Google’s Bard. LLM chatbots are an example of generative AI, in which a program generates new content in response to a prompt based on existing data available to it and patterns within that data; in the case of LLM chatbots, the new content is a conversation with the human user that sounds convincingly like a conversation between 2 humans.

LLM chatbots may prove to have utility to clinical practice and to research. The *New York Times* reported that within 72 hours of the release of ChatGPT, physicians were exploring how it might improve clinical workflow, streamline administrative tasks,

and guide compassionate communication with patients.⁴ Others have suggested that LLM chatbots might be useful in organizing ideas for manuscripts. When Alkaissi and McFarlane provided ChatGPT with PubMed identifiers (PMIDs) and brief notes about each reference, the chatbot was able to assemble “a linguistically coherent text out of the small scattered bullet points, almost like assembling a jigsaw puzzle.”⁵

However, LLM chatbots have raised a number of concerns. The assumptions of “white box” technologies may not be applicable to newer generative AI programs; the processes the programs use are not as transparent even to their programmers, and detecting poor performance is more difficult. A central concern is that LLM chatbot output sounds very convincing, even when the output is flawed or incorrect.

LLM chatbots are echo chambers for the thoughts contained in the data available to them. Their responses range from accurate to wildly inaccurate. Chatbots cannot evaluate the quality or source of the data or of the response, and any bias or misinformation present in the data they use will be reflected in their responses. Their responses may not reflect human ethical reasoning. However, their conversational style can convincingly mimic human interactions and persuade human users that all the information presented is correct, even if it contains errors. There have been several reports of LLM chatbots failing to provide accurate responses, even in cases where the chatbot is instructed to provide references for the response. One lawyer, who used an LLM chatbot to prepare a case, expressed surprise to a judge during a trial when the cases cited in his brief were shown to be nonexistent, and references provided by the chatbot were to irrelevant cases.⁶ Alkaissi and McFarlane⁵ described an example of “artificial hallucination” when they asked ChatGPT to write short essays on a biomedical topic well known to the authors and then fact-checked the responses. They noted that “ChatGPT provided confident responses that seemed faithful and nonsensical when viewed in light of the common knowledge in these areas.” However, when ChatGPT was asked to expand one essay response and to provide references, the authors’ fact checking found that “None of the provided paper titles existed, and all provided PubMed IDs (PMIDs) were of different unrelated papers.”

About the Authors

Cindy L. Munro is coeditor in chief of the *American Journal of Critical Care*. She is dean and professor, School of Nursing and Health Studies, University of Miami, Coral Gables, Florida. **Aluko A. Hope** is coeditor in chief of the *American Journal of Critical Care*. He is an associate professor and physician scientist at Oregon Health and Science University in Portland, Oregon.

“ A measured approach to development and adoption of AI technologies could provide a middle ground that addresses concerns while recognizing the benefits that AI use may yield. ”

De Angelis and colleagues expressed concerns about the potential for an “AI infodemic” resulting from the rise of inaccurate yet convincing health information written by chatbots.⁷ Such an infodemic could result from incorrect information in the chatbot’s data source, from artificial hallucinations, or from malicious manipulation of the prompts given to the chatbot. An AI infodemic could pose serious threats to public health and would be difficult to control or counter.

The use of LLM chatbots in the scientific publication process has been controversial. Guidance from publishers and scientific societies is evolving in response to authors’ use of chatbots in manuscript development, including instances where chatbots were listed among manuscript authors. While some have asked authors to disclose the use of LLM chatbots in preparation of manuscripts (as authors might disclose the use of a statistical software package), others have asserted that chatbots cannot be authors because they cannot meet the International Committee of Medical Journal Editors’ requirement of taking responsibility for the content of a manuscript. Discussions among the publishing team for the American Association of Critical-Care Nurses’ (AACN’s) journals (*American Journal of Critical Care*, *Critical Care Nurse*, *AACN Advanced Critical Care*) about appropriate uses of chatbots are ongoing.

To date, there is no evidence that any AI programs demonstrate “artificial general intelligence,” or the ability to think as humans do. But the emergence of LLM chatbots, whose artificial generative intelligence is so different from the technologies that preceded them, has led to speculation on both the perils and promise of future AI development.^{8,9} Some believe that there is danger of an AI superintelligence emerging, with potential dire consequences for humans; several of the leading experts in AI have called for a moratorium or controls on further development.⁹ Others are hopeful that the capabilities of AI will enhance human abilities and enrich human experience. A measured approach to development and adoption of AI technologies could provide a middle ground that addresses concerns while recognizing the benefits that AI use may yield.

We are not Luddites. We support the appropriate use of technology in critical care clinical practice and research. However, we agree with the National

Institutes of Health’s statement on evidence-based practice that says,

When making decisions on patient care, it is important for health care professionals to use the most relevant, valid, and reliable research in making those decisions. Evidence-based practice is a method of using the best research evidence to strive for optimal patient health outcomes. To identify the best research evidence, you must: (1) Accurately and precisely describe your clinical question, (2) Search high quality resources to find the best information to address your question.¹⁰

At present, no LLM chatbot can be viewed as a high-quality resource, nor can it evaluate the quality of the source material it uses. Both patient care and research are human endeavors. Technology, including LLM chatbots, is a tool rather than a collaborator. Artificial intelligence may be found to be useful in assisting human clinicians and researchers, but it cannot and should not replace us.

The statements and opinions contained in this editorial are solely those of the coeditors in chief.

FINANCIAL DISCLOSURES

None reported.

REFERENCES

1. PubMed Overview. National Center for Biotechnology Information. October 5, 2022. Accessed June 24, 2023. <https://pubmed.ncbi.nlm.nih.gov/about>
2. REDCap. REDCap Consortium. Accessed June 24, 2023. <https://www.project-redcap.org>
3. Roose K. The brilliance and weirdness of ChatGPT. *New York Times*. December 5, 2022. Accessed June 24, 2023. <https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html>
4. Kolata G. When doctors use a chatbot to improve their bedside manner. *New York Times*. June 12, 2023. Accessed June 24, 2023. <https://www.nytimes.com/2023/06/12/health/doctors-chatgpt-artificial-intelligence.html>
5. Alkaiissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. 2023;(2):e35179. doi:10.7759/cureus.35179
6. Weiser B. Here’s what happens when your lawyer uses ChatGPT. *New York Times*. May 27, 2023. Accessed June 24, 2023. <https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>
7. De Angelis L, Baglivo F, Arzilli G, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health*. 2023;11:1166120. doi:10.3389/fpubh.2023.1166120
8. Metz C. How could A.I. destroy humanity? *New York Times*. June 10, 2023. Accessed June 24, 2023. <https://www.nytimes.com/2023/06/10/technology/ai-humanity.html>
9. McCabe D. Microsoft calls for A.I. rules to minimize the technology’s risks. *New York Times*. May 25, 2023. Accessed June 24, 2023. <https://www.nytimes.com/2023/05/25/technology/microsoft-ai-rules-regulation.html>
10. What is evidence-based practice? National Library of Medicine. Accessed June 24, 2023. https://www.nlm.nih.gov/oaet/pubmed/pubmed_in_ebp/01-100.html