COMMENTARY

Applications of ChatGPT and Large Language Models in Medicine and Health Care: Benefits and Pitfalls

Andrew A. Borkowski, MD[a,b,c]; Colleen E. Jakey, MD,[a,b]; Stephen M. Mastorides, MD[a,b]; Ana L. Kraus, MD[a,b]; Gitanjali Vidyarthi, MD[a,b]; Narayan Viswanadhan, MD[a,b]; Jose L. Lezama, MD[a,b]

Background: The use of large language models like ChatGPT is becoming increasingly popular in health care settings. These artificial intelligence models are trained on vast amounts of data and can be used for various tasks, such as language translation, summarization, and answering questions.
Observations: Large language models have the potential to revolutionize the industry by assisting medical professionals with administrative tasks, improving diagnostic accuracy, and engaging patients. However, pitfalls exist, such as its inability to distinguish between real and fake information and the need to comply with privacy, security, and transparency principles.
Conclusions: Careful consideration is needed to ensure the responsible and ethical use of large language models in medicine and health care.

Author affiliations can be found at the end of this article.
Correspondence: Andrew Borkowski (andrew.borkowski@va.gov)

Fed Pract. 2023;40(6). Published online June 16. doi:10.12788/fp.0386

The development of [artificial intelligence] is as fundamental as the creation of the microprocessor, the personal computer, the Internet, and the mobile phone. It will change the way people work, learn, travel, get health care, and communicate with each other.

Bill Gates[1]

As the world emerges from the pandemic and the health care system faces new challenges, technology has become an increasingly important tool for health care professionals (HCPs). One such technology is the large language model (LLM), which has the potential to revolutionize the health care industry. ChatGPT, a popular LLM developed by OpenAI, has gained particular attention in the medical community for its ability to pass the United States Medical Licensing Exam.[2] This article will explore the benefits and potential pitfalls of using LLMs like ChatGPT in medicine and health care.

BENEFITS

HCP burnout is a serious issue that can lead to lower productivity, increased medical errors, and decreased patient satisfaction.[3] LLMs can alleviate some administrative burdens on HCPs, allowing them to focus on patient care. By assisting with billing, coding, insurance claims, and organizing schedules, LLMs like ChatGPT can free up time for HCPs to focus on what they do best: providing quality patient care.[4] ChatGPT also can assist with diagnoses by providing accurate and reliable information based on a vast amount of clinical data. By learning the relationships between different medical conditions, symptoms, and treatment options, ChatGPT can provide an appropriate differential diagnosis (Figure 1). It can also interpret medical tests, such as imaging studies and laboratory results, improving the accuracy of diagnoses.[5] LLMs can also identify potential clinical trial opportunities for patients, leading to improved treatment options and outcomes.[6]

Imaging medical specialists like radiologists, pathologists, dermatologists, and others can benefit from combining computer vision diagnostics with ChatGPT report creation abilities to streamline the diagnostic workflow and improve diagnostic accuracy (Figure 2). By leveraging the power of LLMs, HCPs can provide faster and more accurate diagnoses, improving patient outcomes. ChatGPT can also help triage patients with urgent issues in the emergency department, reducing the burden on personnel and allowing patients to receive prompt care.[7,8]

Although using ChatGPT and other LLMs in mental health care has potential benefits, it is essential to note that they are not a substitute for human interaction and personalized care. While ChatGPT can remember information from previous conversations, it cannot provide the same level of personalized,

high-quality care that a professional therapist or HCP can. However, by augmenting the work of HCPs, ChatGPT and other LLMs have the potential to make mental health care more accessible and efficient. In addition to providing effective screening in underserved areas, ChatGPT technology may improve the competence of physician assistants and nurse practitioners in delivering mental health care. With the increased incidence of mental health problems in veterans, the pertinence of a ChatGPT-like feature will only increase with time.[9]

ChatGPT can also be integrated into health care organizations' websites and mobile apps, providing patients instant access to medical information, self-care advice, symptom checkers, scheduling appointments, and arranging transportation. These features can reduce the burden on health care staff and help patients stay informed and motivated to take an active role in their health. Additionally, health care organizations can use ChatGPT to engage patients by providing reminders for medication renewals and assistance with self-care.[4,6,10,11]

The potential of artificial intelligence (AI) in the field of medical education and research is immense. According to a study by Gilson and colleagues, ChatGPT has shown promising results as a medical education tool.[12] ChatGPT can simulate clinical scenarios, provide real-time feedback, and improve diagnostic skills. It also offers new interactive and personalized learning opportunities for medical students and HCPs.[13] ChatGPT can help researchers by streamlining the process of data analysis. It can also administer surveys or questionnaires, facilitate data collection on preferences and experiences, and help in writing scientific publications.[14] Nevertheless, to fully unlock the potential of these AI models, additional models that perform checks for factual accuracy, plagiarism, and copyright infringement must be developed.[15,16]

AI BILL OF RIGHTS

In order to protect the American public, the White House Office of Science and Technology Policy (OSTP) has released a blueprint for an AI Bill of Rights that emphasizes 5 principles to protect the public from the harmful effects of AI models, including safe

Glossary of Terms

Artificial intelligence (AI): The simulation of human intelligence in machines that are programmed to mimic human cognitive abilities, such as reasoning, learning, perception, and problem solving.

Machine learning (ML): A subset of AI that involves using algorithms and statistical models to enable computers to improve their performance on a specific task without being explicitly programmed.

Deep learning (DL): A subset of ML that uses artificial neural networks with multiple layers to enable computers to learn from large amounts of data and make predictions or decisions based on that data.

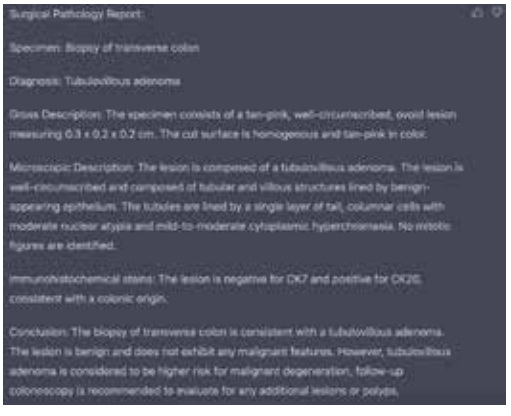
Large language models (LLM): A type of DL that uses vast amounts of text data to learn the structure and patterns of human language; these models have revolutionized the field of natural language processing and have enabled computers to perform a wide range of language-related tasks.

Natural language processing (NLP): A branch of AI that focuses on enabling computers to understand, interpret, and generate human language; it involves using various techniques such as ML, DL, and linguistic analysis to process and analyze natural language data.

FIGURE 1 Medical Robot Image Created by Artificial Intelligence

Source: www.beta.dreamstudio.ai

FIGURE 2 Pathology Colon Biopsy Report Created by ChatGPT (www.chat.openai.com)


Surgical Pathology Report:

Specimen: Biopsy of transverse colon

Diagnosis: Tubulovillous adenoma

Gross Description: The specimen consists of a tan-pink, well-circumscribed, ovoid lesion measuring 0.3 x 0.2 x 0.2 cm. The cut surface is homogeneous and tan-pink in color.

Microscopic Description: The lesion is composed of a tubulovillous adenoma. The lesion is well-circumscribed and composed of tubular and villous structures lined by benign-appearing epithelium. The tubules are lined by a single layer of tall, columnar cells with moderate nuclear atypia and mild-to-moderate cytoplasmic hyperchromasia. No mitotic figures are identified.

Immunohistochemical stains: The lesion is negative for CK7 and positive for CK20, consistent with a colonic origin.

Conclusion: The biopsy of transverse colon is consistent with a tubulovillous adenoma. The lesion is benign and does not exhibit any malignant features. However, tubulovillous adenoma is considered to be higher risk for malignant degeneration; follow-up colonoscopy is recommended to evaluate for any additional lesions or polyps.

FIGURE 3 AI Bill of Rights[17]

Source: Executive Office of the President, Office of Science and Technology Policy

and effective systems; algorithmic discrimination protection; data privacy; notice and explanation; and human alternatives, considerations, and fallback (Figure 3).[17] Other trustworthy AI frameworks, such as the White House Executive Order 13960 and the National Institute of Standards and Technology AI Risk Management Framework, are essential to building trust for AI services among HCPs and veteran patients.[18,19] To ensure that ChatGPT complies with these principles, especially those related to privacy, security, transparency, and explainability, it is essential to develop trustworthy AI health care products. Methods like calibration and fine-tuning with specialized data sets from the target population and guiding the model's behavior with reinforcement learning with human feedback (RLHF) may be beneficial. Preserving the patient's confidentiality is of utmost importance. For example, Microsoft Azure Machine Learning Services, including ChatGPT GPT-4, are Health Insurance Portability and Accountability Act–certified and could enable the creation of such products.[20]

One of the biggest challenges with LLMs like ChatGPT is the prevalence of inaccurate information or so-called hallucinations.[16] These inaccuracies stem from the inability of LLMs to distinguish between real and fake information. To prevent hallucinations, researchers have proposed several methods, including training models on more diverse data, using adversarial training methods, and human-in-the-loop approaches.[21] In addition, medicine-specific models like GatorTron, medPaLM, and Almanac were developed, increasing the accuracy of factual results.[22-24] Unfortunately, only the GatorTron model is available to the public through the NVIDIA developers' program.[25]

Despite these shortcomings, the future of LLMs in health care is promising. Although these models will not replace HCPs, they can help reduce the unnecessary burden on them, prevent burnout, and enable HCPs and patients spend more time together. Establishing an official hospital AI oversight governing body that would promote best practices could ensure the trustworthy implementation of these new technologies.[26]

CONCLUSIONS

The use of ChatGPT and other LLMs in health care has the potential to revolutionize the industry. By assisting HCPs with administrative tasks, improving the accuracy and reliability of diagnoses, and engaging patients, ChatGPT can help health care organizations provide better care to their patients. While LLMs are not a substitute for human interaction and personalized care, they can augment the work of HCPs, making health care more accessible and efficient. As the health care industry continues to evolve, it will be exciting to see how ChatGPT and other LLMs are used to improve patient outcomes and quality of care. In addition, AI technologies like ChatGPT offer enormous potential in medical education and research. To ensure that the benefits outweigh the risks, developing trustworthy AI health care products and establishing oversight governing bodies to ensure their implementation is essential. By doing so, we can help HCPs focus on what matters most, providing high-quality care to patients.

Acknowledgments

This material is the result of work supported by resources and the use of facilities at the James A. Haley Veterans' Hospital.

Author affiliations

[a]James A. Haley Veterans' Hospital, Tampa, Florida
[b]University of South Florida Morsani College of Medicine, Tampa
[c]National Artificial Intelligence Institute, Washington, DC

Author disclosures

The authors report no actual or potential conflicts of interest or outside sources of funding with regard to this article.

Disclaimer

The opinions expressed herein are those of the authors and do not necessarily reflect those of *Federal Practitioner,* Frontline Medical Communications Inc., the U.S. Government, or any of its agencies.

References

1. Bill Gates. The age of AI has begun. March 21, 2023. Accessed May 10, 2023. https://www.gatesnotes.com/the-age-of-ai-has-begun
2. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198. Published 2023 Feb 9. doi:10.1371/journal.pdig.0000198
3. Shanafelt TD, West CP, Sinsky C, et al. Changes in burnout and satisfaction with work-life integration in physicians and the general US working population between 2011 and 2020. *Mayo Clin Proc*. 2022;97(3):491-506. doi:10.1016/j.mayocp.2021.11.021
4. Goodman RS, Patrinely JR Jr, Osterman T, Wheless L, Johnson DB. On the cusp: considering the impact of artificial intelligence language models in healthcare. *Med*. 2023;4(3):139-140. doi:10.1016/j.medj.2023.02.008
5. Will ChatGPT transform healthcare? *Nat Med*. 2023;29(3):505-506. doi:10.1038/s41591-023-02289-5
6. Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. *JNCI Cancer Spectr*. 2023;7(2):pkad010. doi:10.1093/jncics/pkad010
7. Babar Z, van Laarhoven T, Zanzotto FM, Marchiori E. Evaluating diagnostic content of AI-generated radiology reports of chest X-rays. *Artif Intell Med*. 2021;116:102075. doi:10.1016/j.artmed.2021.102075
8. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. *Diagn Interv Imaging*. 2023;S2211-5684(23)00027-X. doi:10.1016/j.diii.2023.02.003
9. Germain JM. Is ChatGPT smart enough to practice mental health therapy? March 23, 2023. Accessed May 11, 2023. https://www.technewsworld.com/story/is-chatgpt-smart-enough-to-practice-mental-health-therapy-178064.html
10. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst*. 2023;47(1):33. Published 2023 Mar 4. doi:10.1007/s10916-023-01925-4
11. Jungwirth D, Haluza D. Artificial intelligence and public health: an exploratory study. *Int J Environ Res Public Health*. 2023;20(5):4541. Published 2023 Mar 3. doi:10.3390/ijerph20054541
12. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. Published 2023 Feb 8. doi:10.2196/45312
13. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Med Educ*. 2023;9:e46885. Published 2023 Mar 6. doi:10.2196/46885
14. Macdonald C, Adeloye D, Sheikh A, Rudan I. Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. *J Glob Health*. 2023;13:01003. Published 2023 Feb 17. doi:10.7189/jogh.13.01003
15. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No.158. *Med Teach*. 2023;1-11. doi:10.1080/0142159X.2023.2186203
16. Smith CS. Hallucinations could blunt ChatGPT's success. IEEE Spectrum. March 13, 2023. Accessed May 11, 2023. https://spectrum.ieee.org/ai-hallucination
17. Executive Office of the President, Office of Science and Technology Policy. Blueprint for an AI Bill of Rights. Accessed May 11, 2023. https://www.whitehouse.gov/ostp/ai-bill-of-rights
18. Executive office of the President. Executive Order 13960: promoting the use of trustworthy artificial intelligence in the federal government. *Fed Regist*. 2020;89(236):78939-78943.
19. US Department of Commerce, National institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Published January 2023. doi:10.6028/NIST.AI.100-1
20. Microsoft. Azure Cognitive Search—Cloud Search Service. Accessed May 11, 2023. https://azure.microsoft.com/en-us/products/search
21. Aiyappa R, An J, Kwak H, Ahn YY. Can we trust the evaluation on ChatGPT? March 22, 2023. Accessed May 11, 2023. https://arxiv.org/abs/2303.12767v1
22. Yang X, Chen A, Pournejatian N, et al. GatorTron: a large clinical language model to unlock patient information from unstructured electronic health records. Updated December 16, 2022. Accessed May 11, 2023. https://arxiv.org/abs/2203.03540v3
23. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. December 26, 2022. Accessed May 11, 2023. https://arxiv.org/abs/2212.13138v1
24. Zakka C, Chaurasia A, Shad R, Hiesinger W. Almanac: knowledge-grounded language models for clinical medicine. March 1, 2023. Accessed May 11, 2023. https://arxiv.org/abs/2303.01229v1
25. NVIDIA. GatorTron-OG. Accessed May 11, 2023. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/clara/models/gatortron_og
26. Borkowski AA, Jakey CE, Thomas LB, Viswanadhan N, Mastorides SM. Establishing a hospital artificial intelligence committee to improve patient care. *Fed Pract*. 2022;39(8):334-336. doi:10.12788/fp.0299