ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 15, Issue, 08, pp.25496-25499, August, 2023
DOI: https://doi.org/10.24941/ijcr.45759.08.2023

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

THE GOOD, THE BAD, AND THE UGLY OF CHAT GPT IN MEDICAL EDUCATION

***Thomas F. Heston MD and Charya Khun MA**

Department of Family Medicine, 1959 NE Pacific Street, Box 356390, Seattle, WA, 98195-6390 USA

ARTICLE INFO

Article History:
Received 14th May, 2023
Received in revised form
18th June, 2023
Accepted 20th July, 2023
Published online 28th August, 2023

Key words:
Generative Language models, Medical Education, ChatGPT, Bard.

**Corresponding Author:*
Thomas F. Heston MD

ABSTRACT

Chat GPT offers interactive and personalized learning, granting students access to vast medical knowledge and potentially enhancing critical thinking and problem-solving skills. However, challenges arise, including misinformation risks, reduced human interaction, and ethical considerations. Future physicians' professional identity and autonomy may be threatened, while over-reliance on Chat GPT can compromise patient safety. Striking a balance is crucial, emphasizing technology integration while preserving humanistic aspects. Mitigation strategies like human oversight, curated content, and critical appraisal skills can address these concerns. Responsible integration empowers future physicians and upholds core medical values, maximizing benefits and preparing students for new complexities introduced by advanced technology in medicine.

Citation: ***Thomas F. Heston MD and Charya Khun MA.*** **2023.** "The Good, the Bad, and the Ugly of Chat GPT in Medical Education*". International Journal of Current Research*, 15, (08), 25496-25499.

INTRODUCTION

Generative language models (GLMs) have witnessed remarkable advancements over the last several decades, driven by a desire to harness the power of computers to communicate with humans with coherent and contextually aware text, as assessed by the Turing Test (Turing, 1950). While neural networks were initially developed in the 1940s, progress towards computers passing the Turing Test began in earnest with the development of recurrent neural networks (RNNs) in the 1980s, then with the long short-term memory (LSTM) model in the 1990s. Then the advent of the transformer model in 2017 revolutionized natural language processing. Although the Turing Test is subjective, there is a growing recognition that modern GLMs can now pass the test and convince most users that they are interacting with another human (Chat, 2023). The concept of artificial neural networks was originally developed in the 1940s (McCulloch, 1943). The back propagation of learning and creating RNNs was a significant step forward in 1986 (Rumelhart, 1986). In 1997, building upon back-propagation, the long short-term memory (LSTM) model was advanced to address the vanishing gradient problem and capture long-term dependencies (Hochreiter, 1997). LSTM was a significant milestone in developing early generative language models (GLMs). Before LSTM, an RNN would do well with the connection between the word immediately preceding the one the RNN was working on but have difficulty putting it in a broader context. While the connection with the immediately preceding word was strong, it was weaker with the second word preceding, and even weaker with the third word preceding. Thus, the strength of the connection (gradient) steadily declined (vanished) the further away a word was from the word the RNN was currently working on. LSTM helped generate new words based not only on the single, immediately preceding word but also on the context created by additional preceding words.

The next significant breakthrough in the development came in 2017 with the development of the transformer model, which revolutionized natural language processing by effectively capturing long-range dependencies using self-attention mechanisms (Ashish, 2017). This used a different model than LSTM; however, the transformer model addressed the same issue of a requirement for long-term context when generating words. This transformer model has heavily influenced the creation of current GLMs. One of the primary GLMs based on the transformer model is the Generative Pre-trained Transformer (GPT) model, developed in 2018 (Radford, 2023). GPT trained neural networks through the unsupervised learning of large amounts of unlabeled data and is the basis for one of the currently dominant GLMs, Chat GPT. Chat GPT is a GLM form of artificial intelligence (AI) that was released to the public by the OpenAI organization on November 30, 2022. Another GLM based on the transformer model is BERT (Bidirectional Encoder Representations from Transformer), initially developed in 2018 and utilizing a bidirectional process to generate words, looking at both the words before and the words following the current word being worked on (Devlin, 2018). Google BARD (Base-Augmented Reader-Decoder) is based on the BERT model. BARD was released to the public on March 21st, 2023. These significant advancements in artificial intelligence have profound implications for medical education, impacting both students and educators. Furthermore, society holds a vested interest in the implementation of these tools in healthcare. While the transformative benefits are widely acknowledged, there is also a growing recognition of potential risks associated with AI's rapid and substantial progress.

The Good: Improved Learning

GLMs like Chat GPT and Google Bard offer numerous benefits in medical education. They revolutionize learning by engaging students

in conversational language, providing round-the-clock availability. With GLMs, students can participate in dynamic one-on-one conversations, pose academic questions, role-play to enhance interview skills, and receive guidance for personal issues such as scheduling and planning (Heston, 2023). Clinical reasoning of artificial chat bots is already reasonably high, showing great potential to assist in all aspects of medical education (Gilson, 2023; Strong , 2023). In contrast, traditional classroom settings often limit students' opportunities to question the teacher, leaving any inquiries unaddressed. Before GLMs, students with questions outside of classroom hours typically relied on textbooks or internet searches, which consumed considerable time. GLMs, however, are pre-trained on vast databases, allowing them to offer near-instantaneous responses to student queries. This significantly reduces search time and enables students to learn more about the topics they wish to explore. Rather than following a rigid learning path dictated by the teacher during a lecture, students can now take detours, satisfying their curiosity and engaging more thoroughly with the subject matter. The ability of GLMs to respond directly to a wide variety of questions fosters active learning and empowers students to navigate complex medical concepts in a personalized and engaging manner. Here are a few specific examples of how GLMs can assist in medical education:

- Personalized Learning: GLMs adapt to individual learners, tailoring content and feedback based on specific needs, goals, and learning styles (Abd-Alrazaq, 2023). By analyzing students' interactions and performance, these models provide targeted recommendations, identify areas for improvement, and offer personalized educational interventions. This customized approach enhances students' ability to grasp and retain medical knowledge effectively.
- Simulating Patient Scenarios: GLMs simulate case studies, stimulating critical thinking and problem-solving skills. Through interactive conversations, students analyze information, evaluate options, and make informed decisions, mirroring real-world medical practice (Karabacak, 2023). These simulations nurture analytical thinking abilities and prepare students for the complexities they may encounter in their future careers.
- Time Management Support: GLMs can assist students in managing their time effectively. By providing guidance and reminders for scheduling, planning study sessions, and organizing tasks, these models help students stay on track and optimize their learning efficiency. GLMs can also help students prioritize their workload, ensuring they allocate sufficient time to different subjects or assignments.
- Differential Diagnoses: The GLM can help create differential diagnoses. Medical students can input complex and unusual symptoms into Chat GPT and prompt it to generate a list of possible diseases. This is especially useful for identifying rare syndromes and illnesses often overlooked in medical school. However, it is important to note that users should never share personal patient information with Chat GPT.
- Networking: The GLM can assist with improving a student's social network. Networking is vital in medical school because it allows students to gain mentors, internships, research opportunities, resources, and exposure to different medical specialties. Students can use Chat GPT to draft cold emails and LinkedIn messages to physicians and ask them for mentorship or research opportunities, significantly mitigating the anxiety that comes from reaching out to an unfamiliar person. Students can use the GLM to draft comments and responses on platforms such as Instagram, Twitter, or Reddit to help build connections and have meaningful conversations with people within their specialty of interest. Furthermore, Chat GPT can be used to prepare questions for conference speakers and webinars so that their questions and comments stand out to a prominent speaker.

Overall, GLMs offer versatile support for medical education by providing personalized learning experiences, simulating patient scenarios, and aiding students in managing their time effectively. By harnessing the capabilities of these models, educators and learners can enhance the learning process, deepen understanding, and cultivate essential skills for future medical practitioners.

The Bad: Misinformation

GLMs, such as Chat GPT and Google Bard, have made notable strides in medical education. However, it is crucial to acknowledge the potential problems and risks that accompany their use. One significant concern is the possibility of misinformation, as GLMs may inadvertently provide students with incorrect or poorly contextualized information, leading to unintended consequences. Firstly, students may unknowingly receive wrong information from GLMs. These models are trained on extensive databases encompassing many sources, some of which may contain outdated or inaccurate data. For instance, if a student asks a GLM about the latest treatment guidelines for a particular condition, the model's response may be outdated since its training concluded in 2021 (Chat GPT 3.5). This can lead students to unknowingly acquire obsolete medical knowledge that may adversely affect patient care. Secondly, GLMs face challenges in discerning the reliability and validity of information from the internet (Májovský, 2023). Although they can access online resources to answer queries, distinguishing between credible and unreliable sources remains a limitation. Consequently, students may unknowingly receive inaccurate or misleading information from GLMs, as these models cannot effectively evaluate the credibility of online content. This can significantly affect students' understanding and application of medical concepts. Furthermore, the absence of human teachers in the learning process amplifies the risk of students acquiring erroneous knowledge. GLMs lack the nuanced understanding and contextualization that human educators provide. Consequently, students may develop a flawed understanding of medical concepts, leading to potential harm in real-world patient care situations. Without human oversight, students may fail to grasp medical knowledge's subtleties, complexities, and evolving nature, resulting in detrimental consequences for patients and students.

Finally, GLMs are not immune to biases, which can hurt medical education. These models are trained on large datasets that can reflect inherent biases present in the data, such as disparities in healthcare access or representation of certain demographics. Consequently, the responses provided by GLMs may inadvertently perpetuate or amplify these biases, influencing students' understanding of medical concepts. If students rely solely on GLMs for information, they may be exposed to a limited or skewed perspective, hindering their ability to develop a comprehensive and unbiased understanding of healthcare. It is crucial to be mindful of the potential biases within GLMs and supplement their use with critical thinking, diverse perspectives, and human guidance to ensure a more equitable and inclusive medical education.

The Ugly: Ethical Dangers

GLMs, such as ChatGPT and Google Bard, have introduced ethical challenges and concerns in medical education. While they offer numerous benefits, addressing the potential ethical challenges and implications associated with their use is important. Firstly, the accessibility and conversational nature of GLMs raise concerns about academic integrity and the potential for cheating. Students can exploit these models by seeking direct answers to exam questions or relying heavily on them for assignments without actively engaging in critical thinking and independent learning. For instance, a student could submit a paper written predominantly by a GLM, plagiarizing the work and undermining the development of their own knowledge and skills (Dwivedi, 2023). Secondly, GLMs can potentially isolate students from one another and their teachers. The convenience and availability of GLMs may tempt students to spend more time interacting with the models and less time engaging with their peers and instructors in face-to-face interactions. This shift can hinder the development of crucial people skills, such as effective communication, collaboration, and teamwork, which are vital in the medical field. By relying too heavily on GLMs, students may miss out on the social and interactive learning experiences essential for their professional growth.

Furthermore, the abundance of data and information provided by GLMs can have unintended consequences on learning outcomes. While access to vast amounts of data can be beneficial, it can also create a dependency on the models, potentially reducing students' motivation to seek out and critically evaluate information independently. Students may become intellectually dependent on the GLMs, leading to a decrease in deep learning and the ability to analyze, synthesize, and apply knowledge. This over-reliance on GLMs may hinder students' development of essential critical thinking and problem-solving skills, limiting their ability to think creatively and adapt to complex medical scenarios.

The Hopeful: Strategies and Recommendations

GLMs have become an integral part of medical education, and it is evident that they are here to stay. Rather than shying away from these tools, embracing them and exploring strategies to maximize their benefits, minimize potential drawbacks, and address ethical concerns is essential. By proactively guiding the use of GLMs and incorporating effective teaching practices, we can harness their potential to transform medical education. To begin with, the integration of GLMs should be guided by educators rather than leaving students to navigate this new technology on their own. Teachers are crucial in setting expectations, providing guidance, and curating appropriate resources. By actively engaging with GLMs and staying informed about their capabilities and limitations, educators can effectively incorporate these tools into the curriculum, ensuring optimal use for enhancing student learning. Furthermore, addressing the issue of cheating is essential in maintaining academic integrity. While it is true that cheating existed before the advent of GLMs, it remains a concern. To mitigate this, implementing honor codes and emphasizing ethical conduct can help create a culture of integrity. Clear communication regarding the expectations, consequences of academic misconduct, and the importance of independent learning is vital in deterring cheating and promoting a sense of responsibility among students.

Promoting small-group, interactive learning can be a powerful strategy to maximize the benefits of GLMs. Students are encouraged to collaborate, engage in discussions, and share diverse perspectives by requiring attendance and active participation in these sessions. This approach fosters critical thinking, communication skills, and teamwork – all vital aspects of a successful medical career. Students can develop a well-rounded skill set by complementing GLM interactions with face-to-face interactions, incorporating technological and interpersonal abilities. Another strategy to consider is minimizing the reliance on traditional lectures. While lectures have their place in education, they should be balanced with interactive and engaging learning experiences. Utilizing GLMs for self-paced learning and incorporating flipped classroom models can create opportunities for deeper exploration, problem-solving, and interactive discussions. Students can comprehensively understand medical concepts by reducing passive learning and encouraging active engagement. In addition to the mentioned strategies, it is important to continuously explore new ways to leverage GLMs effectively in medical education. This includes refining the training of these models to address biases, promoting inclusivity and cultural sensitivity, and ensuring that GLMs provide accurate and up-to-date information. Collaborative efforts between educators, researchers, and technology developers can help refine and improve the capabilities of GLMs, leading to more reliable and effective educational tools.

CONCLUSION

Integrating generative language models (GLMs) in medical education brings opportunities and challenges. Educators play a pivotal role in maximizing the benefits of GLMs by providing guidance, curating resources, and integrating these models into the curriculum. Maintaining academic integrity is crucial, employing strategies like honor codes to minimize cheating. Promoting small-group, interactive learning fosters essential people skills while minimizing reliance on traditional lectures. Continued refinement, addressing biases, promoting inclusivity, and ensuring accurate information are ongoing priorities. Through careful implementation, improvement, and a commitment to ethical practices, GLMs can advance medical education, preparing future healthcare professionals for the complexities of patient care.

Key Points

- Chat GPT offers personalized learning, interactive patient simulations, and time management support. However, risks exist, including misinformation, reduced human interaction, and ethical concerns like cheating.
- Strategies to maximize the benefits of Chat GPT while minimizing risks include supervision by educators, small group learning, and refining training to address biases.
- Responsible integration of Chat GPT in medical education can enhance learning and prepare students, but maintaining academic integrity, critical thinking skills, and humanistic aspects of medicine is crucial.

Conflict of Interests: No competing interests.

Funding Information: Self-funded, no external funding.

Glossary of Abbreviations

AI: artificial intelligence

BARD: Base-Augmented Reader-Decoder

BERT: Bidirectional Encoder Representations from Transformet

GLM: Generative language model

GPT: Generative pre-trained transformer

LSTM: Long short-term memory

RNN: Recurrent neural network long short-term memory

RNN: Recurrent neural network

REFERENCES

Turing AM. Mind. Mind. 1950;59(236):433–60.

ChatGPT Passes Turing Test: A Turning Point for Language Models [Internet]. [cited 2023 Jul 10]. Available from: https://www.mlyearning.org/chatgpt-passes-turing-test/

McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943 Dec;5(4):115–33. DOI: 10.1007/BF02478259.

Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533–6. DOI: 10.1038/323533a0.

Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. DOI: 10.1162/neco.1997.9.8.1735. PMID: 9377276.

Ashish V. Attention is all you need. Advances in neural information processing systems. 2017;30.

Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. [Internet]. 2018 [cited 2023 Jun 20]. Available from: https://web.archive.org/web/20230622213848/https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv. 2018; DOI: 10.48550/arxiv.1810.04805.

Heston T. Prompt Engineering: for Students of Medicine and Their Teachers. Independently published; 2023.

Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, 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 Feb 8;9:e45312. DOI: 10.2196/45312. PMID: 36753318. PMCID: PMC9947764.

Strong E, DiGiammarino A, Weng Y, Basaviah P, Hosamani P, Kumar A, et al. Performance of ChatGPT on free-response, clinical reasoning exams. medRxiv. 2023 Mar 29; DOI: 10.1101/2023.03.24.23287731. PMID: 37034742. PMCID: PMC10081420.

Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ. 2023 Jun 1;9:e48291. DOI: 10.2196/48291. PMID: 37261894. PMCID: PMC10273039.

Karabacak M, Ozkara BB, Margetis K, Wintermark M, Bisdas S. The advent of generative language models in medical education. JMIR Med Educ. 2023 Jun 6;9:e48163. DOI: 10.2196/48163. PMID: 37279048. PMCID: PMC10282912.

Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. J Med Internet Res. 2023 May 31;25:e46924. DOI: 10.2196/46924. PMID: 37256685. PMCID: PMC10267787.

Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manage. 2023 Aug;71:102642. DOI: 10.1016/j.ijinfomgt.2023.102642.

****