Large Language Models and User Trust: Focus on Healthcare

Avishek Choudhury, Zaira Chaudhry

Submitted to: Journal of Medical Internet Research
on: January 25, 2024

Table of Contents

Large Language Models and User Trust: Focus on Healthcare

Avishek Choudhury[1] PhD; Zaira Chaudhry[1] MD

[1]Industrial and Management Systems Engineering West Virginia University Morgantown US

Corresponding Author:
Avishek Choudhury PhD
Industrial and Management Systems Engineering
West Virginia University
321 Engineering Sciences Bdlg
1306 Evansdale Drive
Morgantown
US

Abstract

The role of artificial intelligence (AI), particularly large language models (LLM) like Chat Generative Pre-Trained Transformer (ChatGPT), has garnered significant attention in healthcare. This paper focuses on how user expertise and their trust in the technology can influence LLMs' effectiveness in healthcare. Our arguments instigate the following questions: are we and our healthcare system ready to integrate LLMs? If yes, is there a policy explicitly stating in what capacity it could be used to reduce clinical workload before its dissemination? Will the ease of generating content with AI stifle the development of creativity and critical thinking in medical students accustomed to technology providing immediate solutions? Additionally, we elucidate risk factors such as the possibility of a self-referential loop and accountability problems emerging due to LLMs in healthcare. While these problems have yet to materialize, they represent a likely challenge as LLMs advance and proliferate in healthcare. A thoughtful, deliberate approach to integrating LLMs into healthcare can mitigate risks associated with overreliance and deskilling, ensuring that it complements rather than compromises the quality of care. By leveraging AI's strengths and compensating for its limitations through human oversight, healthcare can harness the benefits of this technology to improve outcomes, enhance patient care, and support healthcare professionals in their vital work. Thus, the path forward involves embracing generative AI's potential while remaining vigilant about its limitations, ensuring that its integration enhances rather than diminishes the human element in healthcare.

(JMIR Preprints 25/01/2024:56764)
DOI: https://doi.org/10.2196/preprints.56764

Original Manuscript

Large Language Models and User Trust: Focus on Healthcare

Avishek Choudhury (Ph.D.)

avishek.choudhury@mail.wvu.edu

Corresponding

Avishek Choudhury (PhD)

Industrial and Management Safety Engineering

West Virginia University

Large Language Models and User Trust: Focus on Healthcare

Abstract

The role of artificial intelligence (AI), particularly large language models (LLM) like Chat Generative Pre-Trained Transformer (ChatGPT), has garnered significant attention in healthcare. This paper focuses on how user expertise and their trust in the technology can influence LLMs' effectiveness in healthcare. Our arguments instigate the following questions: are we and our healthcare system ready to integrate LLMs? If yes, is there a policy explicitly stating in what capacity it could be used to reduce clinical workload before its dissemination? Will the ease of generating content with AI stifle the development of creativity and critical thinking in medical students accustomed to technology providing immediate solutions? Additionally, we elucidate risk factors such as the possibility of a self-referential loop and accountability problems emerging due to LLMs in healthcare. While these problems have yet to materialize, they represent a likely challenge as LLMs advance and proliferate in healthcare. A thoughtful, deliberate approach to integrating LLMs into healthcare can mitigate risks associated with overreliance and deskilling, ensuring that it complements rather than compromises the quality of care. By leveraging AI's strengths and compensating for its limitations through human oversight, healthcare can harness the benefits of this

technology to improve outcomes, enhance patient care, and support healthcare professionals in their vital work. Thus, the path forward involves embracing generative AI's potential while remaining vigilant about its limitations, ensuring that its integration enhances rather than diminishes the human element in healthcare.

INTRODUCTION

The role of artificial intelligence (AI), particularly large language models (LLM) like Chat Generative Pre-Trained Transformer (ChatGPT), has garnered significant attention in healthcare. Fundamentally LLMs operates by learning from vast datasets derived from human intellectual outputs, such as books, articles, and other forms of digital communication. By continuously learning from an expansive array of real-world data, these models refine their ability to produce contextually relevant outputs. LLMs have shown promise in enhancing interactions between providers and patients, streamlining the creation of patient education materials, and aiding in complex medical assessments and management [1-5]. Additionally, their applications extend to scientific writing, e-healthcare, and patient classification via electronic medical records [6-8]. They have also been used for task-oriented dialogue, health and fitness coaching, and enhancing productivity and learning efficiency in higher education [9, 10]. As these technologies become more ingrained in our daily lives, the potential for transformative effects on society grows, marking a significant shift in paradigms.

The full integration of existing AI models into healthcare—a field where the trust in AI is crucial due to the significant impact of decision-making—is still a work in progress [11]. At the same time, efforts to develop standardized protocols for the deployment of AI in healthcare are underway, yet they have not reached a point of completion [12]. This endeavor is critical for ensuring AI's safe and effective use in healthcare settings. Additionally, the challenge of evaluating AI in healthcare is exacerbated by a lack of comprehensive and standardized metrics [13]. This void is something that

researchers and policymakers are actively working to address by creating robust evaluation frameworks that could be applied universally. The regulatory landscape for healthcare AI is evolving, with a growing agreement on the necessity for policies that address ethical considerations, data privacy, and patient safety, alongside frameworks that hold AI systems and their developers accountable for the outcomes of their use in patient care [11].

Despite these ongoing challenges and developments, LLMs are already being deployed in the public sphere, utilized by healthcare workers, researchers, and the public for a variety of healthcare-related tasks. This early adoption highlights the critical need for continued discourse and development in this area to ensure that AI's integration into healthcare is both safe and effective. We acknowledge that LLMs can expedite clinical tasks and potentially allow clinicians more time for patient interaction, these models are not without limitations [14-18].

In this paper we focus on how user expertise and their trust in the technology can influence LLMs effectiveness in healthcare. While the users of LLMs can be diverse, including patients and insurance providers, the focus of this paper is limited to the domain experts in healthcare, i.e., clinicians. We acknowledge that trust relationships could significantly differ for non-healthcare users. Additionally, we also elucidate risk factors such as the possibility of self-referential loop in LLMs learning, deskilling of users, and accountability problems emerging due to LLMs in healthcare. While these problems have not yet materialized, it represents a likely challenge as LLMs continues to advance and proliferate in healthcare. We recognize that a discussion surrounding the incorporation of LLMs like ChatGPT into healthcare is essential, and it's important to note that this paper does not stem from a resistance to technological progress. Rather, the critical viewpoint on LLMs originates from a position of caution and a sense of duty to ensure their safe integration and adoption.

USER EXPERTISE AND TRUST IN LLMS

The integration of LLMs like ChatGPT into healthcare, presents a nuanced landscape of

opportunities and challenges that hinge significantly on user expertise and trust. In this section we discuss the critical need for a balanced approach towards adopting LLMs, emphasizing the importance of user ability to detect errors and the willingness to verify AI-generated content, thereby ensuring the trust in technology augments human work without introducing undue risks.

User expertise: ability to detect LLM error

The use of LLMs, particularly models like ChatGPT, presents a range of possibilities and challenges that vary depending on the user's expertise and intent, delineating into two primary user categories—subject matter experts and those seeking assistance due to a lack of knowledge.

Subject matter experts may employ ChatGPT to handle routine, time-consuming tasks, enabling them to allocate more time to complex or urgent issues. For instance, a doctor might use ChatGPT to quickly summarize patient notes or streamline administrative tasks without compromising on the quality of patient care. These users have the advantage of being able to critically evaluate the AI's output, verify its accuracy (deviation from clinical standards), and make necessary corrections. The expertise of such users acts as a safeguard against potential errors, ensuring that the AI's assistance enhances productivity without introducing risk.

On the other hand, individuals who turn to ChatGPT due to a lack of knowledge or expertise in a particular area face a different set of challenges [19]. When a patient uses ChatGPT for self-diagnosis or a researcher employs it to draft a manuscript on an unfamiliar medical subspecialty, the potential for misinformation or inaccuracies becomes a significant concern. These users may not possess the necessary knowledge to critically assess the AI's output [20]. The ability of chatGPT to generate fake but persuasive response further exacerbate the risks making users vulnerable to accepting erroneous information as fact [21]. The crux of the problem lies in the user's ability to verify the accuracy and relevance of the AI-generated content.

User trust: willingness to review LLM output

Alike user ability to detect potential AI errors, their willingness to scrutinize them is equally important. A user may have the ability and necessary expertise but may not be willing to review LLM generated outcome due to factors including trust in the technology and workload.

In the context of user trust in LLMs or ChatGPT, for instance, will largely determine whether they will accept or question AI results. In other words, if a doctor has high trust (blind trust) in the LLM, they might be more inclined to accept the AI's suggestion without extensive further verification [22, 23]. This can occur when the system demonstrates high accuracy and reliability over a certain period, leading users to believe in its infallibility. The danger lies in the potential for users to become complacent, foregoing the critical step of verifying the AI's responses. This blind trust might persist unchallenged until a point of failure or harm, which can have serious implications, especially in sensitive and high-stakes environments like that of healthcare. Conversely, if the provider harbors doubts about the LLM's reliability or has experienced or heard of inaccurate diagnoses from similar AI tools in the past, they might question the AI-generated diagnosis. This skepticism could lead them to seek additional tests or second opinions before making a decision, potentially delaying treatment but also possibly preventing a misdiagnosis.

Given LLM’s inherent reliance on the quality and scope of its training data, there is a tangible risk of the system generating and propagating false or low-quality information [24]. This issue is not merely theoretical but a practical concern that has manifested in various instances [25]. In high-risk sectors like healthcare, where decisions based on accurate and reliable information are paramount, the consequences of misinformation can be particularly grave.

Therefore, the solution is often encouraging doctors (users) to verify AI outcomes. The pivotal consideration here is whether the verification of AI outcomes by healthcare staff negates the purported reduction in workload. If healthcare professionals are required to meticulously check each AI-generated output for accuracy, the time saved through automation may be offset by the time spent

on verification.

Approaching AI integration in healthcare with a critical mindset is important. It is crucial for users to have a clear understanding of a technology's actual performance, distinguishing it from the exaggerated expectations set by media hype. These risks underscore the importance of asking the question: are we and our healthcare system ready to integrate LLMs? If yes, is there a policy in place explicitly stating in what capacity it could be used to reduce clinical workload before its dissemination?

FUTURE RISK CONSIDERATIONS

In this section we will discuss the possibility of a self-referential loop in LLM learning, deskilling of users, and accountability problems emerging from LLMs in healthcare. Note that these problems have not yet materialized, it represents a likely challenge as LLMs continues to advance and proliferate in society.

LLM paradox of self-referential loop (learning from itself)

In a scenario where LLMs become widely adopted in the healthcare industry for tasks like manuscript writing, educational material creation, clinical text summarization, and risk identification, the possibility of a *self-referential loop* does emerge as a significant concern. This paradox occurs when AI-generated human-like content becomes so widespread that the AI begins to reference its own generated content, potentially leading to an echo chamber effect where original, human-generated insights become diluted or harder to distinguish from AI-generated content. While this problem of a self-referential loop in AI-generated content, particularly in the healthcare industry, has not yet materialized, it represents a likely challenge as generative AI continues to advance and proliferate.

The consequence of a self-referential loop in LLMs can lead to several problematic outcomes, including the propagation of biases [26], increased homogeneity in generated data, and ultimately, hindered performance. AI systems learn from the data they are fed, and if these data include biases, the AI is likely to replicate and even amplify these biases in its outputs [27]. In a self-referential loop, the problem becomes compounded. As the AI references its own biased outputs to generate new content, these biases can become more entrenched, making them harder to identify and correct.

Addressing this challenge requires a nuanced approach to developing and integrating LLMs technologies into societal frameworks. It involves fostering a symbiotic relationship between human intellect and LLM capabilities, ensuring that AI serves as a tool for augmenting human intellect rather than replacing it. Strategies for maintaining the diversity and quality of training data, including the deliberate inclusion of varied and novel human-generated content, will be critical.

Risk of deskilling

As individuals come to rely more on LLMs for routine tasks, such as the synthesis of patient information or the interpretation of medical data, there is a possibility that their skills in these critical areas may diminish over time due to reduced practice [28]. This situation is compounded by the AI's ability to quickly furnish answers to medical inquiries, which might decrease the motivation for in-depth research and learning, consequently affecting the professionals' knowledge depth and critical thinking capabilities.

It is crucial to note that the discussion here does not assert that LLMs will definitively lead to the deskilling of current practitioners in the healthcare sector. These professionals have developed their expertise through extensive experience and rigorous academic training, establishing a solid foundation that is not readily compromised by the integration of AI tools. Instead, the concern is more pronounced for the next generation of healthcare professionals, particularly medical students who might increasingly utilize AI for educational tasks and learning activities where over delegating

tasks to AI could attenuate the development of critical analytical skills and a comprehensive understanding of medical concepts, traditionally cultivated through deep engagement with the material [28, 29]. The critical question emerges: *will the ease of generating content with AI stifle the development of creativity and critical thinking in younger generations accustomed to technology providing immediate solutions?*

If future generations of clinicians grow accustomed to AI doing the bulk of diagnostic review and analysis, there is a risk that their own diagnostic skills might not develop as fully. More critically, should they be required to review patient charts manually—due to AI failures—they may find the task daunting, or lack the detailed insight that manual review processes help to cultivate. The crux of the issue lies in ensuring that reliance on technology should not come at the expense of fundamental skills and knowledge. The challenge is to ensure that the deployment of AI technologies complements human abilities without diminishing the need for critical thinking, reasoning, and creativity.

LLM accountability

The risks of LLM or generative AI integration in to high risk organizations like healthcare has also been noted in The Algorithmic Accountability Act of 2023 and H.R.3369 - Artificial Intelligence Accountability Act [30, 31]. The acts represent a critical legislative step towards ensuring the responsible use of algorithms. The act calls for the creation of standardized procedures and assessment frameworks to evaluate the effectiveness and consequences of these systems, reflecting an understanding of the complex ethical and regulatory challenges posed by AI in decision-making processes, particularly in healthcare. The act is in dialogue with the wider conversation on the ethics of AI, advocating for an approach that emphasizes response-ability—the capacity to respond ethically to the challenges posed by algorithmic decision-making. This perspective is crucial for

developing impact assessments and frameworks aimed at promoting fairness and preventing discriminatory practices within algorithmic systems.

The implications of this act on the integration of large language models (LLMs) in healthcare is profound. It necessitates that developers and users of LLMs in healthcare contexts rigorously evaluate and demonstrate the ethical use of these technologies. Healthcare providers using AI must ensure that their systems are transparent, their decision-making processes are understandable, and that they have measures in place to mitigate bias and protect patient data. This legislative framework could lead to the enhancement of trust in AI applications within healthcare by providing clear guidelines for their development and use, thus fostering an environment where AI can be harnessed to improve patient outcomes while safeguarding ethical standards and patient rights.

Conclusion

It's important to acknowledge that performance of LLMs like ChatGPT today does not guarantee their performance tomorrow. This discourse around its integration into healthcare is not rooted in opposition to technological advancement, rather from a place of caution and responsibility.

LLMs has the potential to be a substantial boon to the healthcare industry, offering to streamline workflows, enhance the accuracy of patient data processing, and even support diagnostic and treatment planning processes. Its value, however, is contingent upon a systematic and informed integration into healthcare systems. Recognizing that LLMs, like any technology, is fallible is crucial to its successful adoption. Its performance is temporal and will changes as new data is fed to its algorithm. This acknowledgment underpins the necessity for robust oversight mechanisms, ongoing evaluation of AI-driven outputs for accuracy and relevance, and clear guidelines on its role as an assistive tool rather than a standalone decision-maker.

A thoughtful, deliberate approach to integrating generative AI into healthcare can mitigate risks

associated with overreliance and deskilling, ensuring that it complements rather than compromises the quality of care. By leveraging AI's strengths and compensating for its limitations through human oversight, healthcare can harness the benefits of this technology to improve outcomes, enhance patient care, and support healthcare professionals in their vital work. Thus, the path forward involves embracing generative AI's potential while remaining vigilant about its limitations, ensuring that its integration enhances rather than diminishes the human element in healthcare.

References

1. Sedaghat S. Early Applications of ChatGPT in Medical Practice, Education and Research. Clinical Medicine. 2023. doi: 10.7861/clinmed.2023-0078.

2. Hung Y-C, Chaker SC, Sigel M, Saad M, Slater ED. Comparison of Patient Education Materials Generated by Chat Generative Pre-Trained Transformer Versus Experts. Annals of Plastic Surgery. 2023. doi: 10.1097/sap.0000000000003634.

3. Galido PV, Butala S, Chakerian M, Agustines D. A Case Study Demonstrating Applications of ChatGPT in the Clinical Management of Treatment-Resistant Schizophrenia

 &#X0D. Cureus. 2023. doi: 10.7759/cureus.38166.

4. Morrison M, Nobles V, Johnson-Agbakwu C, Bailey CV, Liu L. Classifying Refugee Status Using Common Features in EMR**. Chemistry & Biodiversity. 2022. doi: 10.1002/cbdv.202200651.

5. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the Performance of ChatGPT in Answering Questions Regarding Cirrhosis and Hepatocellular Carcinoma. 2023. doi: 10.1101/2023.02.06.23285449.

6. Chen TJ. ChatGPT and Other Artificial Intelligence Applications Speed Up Scientific Writing. Journal of the Chinese Medical Association. 2023. doi: 10.1097/jcma.0000000000000900.

7. Malamas N, Papangelou K, Symeonidis AL. Upon Improving the Performance of Localized

Healthcare Virtual Assistants. Healthcare. 2022. doi: 10.3390/healthcare10010099.

8. Wang DQ, Feng L, Ye J, Zou JH, Zheng Y. Accelerating the Integration of ChatGPT and Other Large-scale AI Models Into Biomedical Research and Healthcare. Medcomm – Future Medicine. 2023. doi: 10.1002/mef2.43.

9. Vardhan M, Hegde N, Nathani D, Rosenzweig E, Karthikesalingam A, Seneviratne M. Infusing Behavior Science Into Large Language Models for Activity Coaching. 2023. doi: 10.1101/2023.03.31.23287995.

10. Firaina R, Sulisworo D. Exploring the Usage of ChatGPT in Higher Education: Frequency and Impact on Productivity. Buletin Edukasi Indonesia. 2023. doi: 10.56741/bei.v2i01.310.

11. Harris LA. Artificial Intelligence: Background, Selected Issues, and Policy Considerations. 2021.

12. DoC. AI Accountability Policy Request for Comment. 2023.

13. Shinners L, Aggar C, Grace S, Smith S. Exploring Healthcare Professionals' Understanding and Experiences of Artificial Intelligence Technology Use in the Delivery of Healthcare: An Integrative Review. Health Informatics Journal. 2019. doi: 10.1177/1460458219874641.

14. Hoffmann J, Borgeaud S, Arthur M, Buchatskaya E, Cai T, Rutherford E, et al. Training Compute-Optimal Large Language Models. 2022. doi: 10.48550/arxiv.2203.15556.

15. Ghafoor S, Yahya A. "Assessing ChatGPT's Performance in Answering Cervical Cancer Screening Questions to ChatGPT-generated Clinical Vignettes: A Pilot Study". 2023. doi: 10.21203/rs.3.rs-3152561/v1.

16. Seth I, Kenney P, Bulloch G, Hunter-Smith DJ, Thomsen JB, Rozen WM. Artificial or Augmented Authorship? A Conversation With a Chatbot on Base of Thumb Arthritis. Plastic and Reconstructive Surgery Global Open. 2023. doi: 10.1097/gox.0000000000004999.

17. Yeung JA, Kraljević Ž, Luintel A, Balston A, Idowu E, Dobson R, Teo J. AI Chatbots Not Yet Ready for Clinical Use. 2023. doi: 10.1101/2023.03.02.23286705.

18. Denejour D. ChatGPT in Pharmacometrics? Potential Opportunities and Limitations. 2023. doi: 10.22541/au.168245333.32388275/v1.

19. 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. International Journal of Information Management. 2023;71:102642.

20. Dempere J, Modugu K, Hesham A, Ramasamy LK. The impact of ChatGPT on higher education. Frontiers in Education. 2023 2023-September-08;8. doi: 10.3389/feduc.2023.1206936.

21. Shahsavar Y, Choudhury A. User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. JMIR Hum Factors. 2023 2023/5/17;10:e47564. doi: 10.2196/47564.

22. Choudhury A, Shamszare H. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. J Med Internet Res. 2023 2023/6/14;25:e47184. doi: 10.2196/47184.

23. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res. 2020 2020/6/19;22(6):e15154. doi: 10.2196/15154.

24. De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, Rizzo C. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. Frontiers in Public Health. 2023 2023-April-25;11. doi: 10.3389/fpubh.2023.1166120.

25. Gravel J, D’Amours-Gravel M, Osmanlliu E. Learning to fake it: limited responses and fabricated references provided by ChatGPT for medical questions. Mayo Clinic Proceedings: Digital Health. 2023;1(3):226-34.

26. Meyer JG, Urbanowicz RJ, Martin PCN, O’Connor K, Li R, Peng P-C, et al. ChatGPT and large language models in academia: opportunities and challenges. BioData Mining. 2023 2023/07/13;16(1):20. doi: 10.1186/s13040-023-00339-9.

27. Nazir A, Wang Z. A comprehensive survey of ChatGPT: Advancements, applications,

prospects, and challenges. Meta-Radiology. 2023 2023/09/01/;1(2):100022. doi: https://doi.org/10.1016/j.metrad.2023.100022.

28. Lam K. ChatGPT for low- and middle-income countries: a Greek gift? The Lancet Regional Health – Western Pacific. 2023;41. doi: 10.1016/j.lanwpc.2023.100906.

29. Lo CK. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. Education Sciences [Internet]. 2023; 13(4).

30. H.R.3369 — 118th Congress (2023-2024). Congress; 2023 [cited 2024]; Available from: https://www.congress.gov/bill/118th-congress/house-bill/3369.

31. Algorithmic Accountability Act of 2023. Office of U.S. Senator Ron Wyden; 2023 [cited 2024]; Available from: https://www.wyden.senate.gov/imo/media/doc/algorithmic_accountability_act_text.pdf.