Exploring patient trust in clinical advice from AI-driven LLMs like ChatGPT for self-diagnosis

Delong Du[1], Richard Paluch[1], Gunnar Stevens[1], Claudia Müller[1]

[1] University of Siegen, Germany
[a] Corresponding author; E-Mail: delong.du@uni-siegen.de

Trustworthy clinical advice is crucial but burdensome when seeking health support from professionals. Inaccessibility and financial burdens present obstacles to obtaining professional clinical advice, even when healthcare is available (Taber et al., 2015). Consequently, individuals often resort to self-diagnosis, utilizing medical materials to validate the health conditions of themselves, their families, and friends. However, the convenient method of self-diagnosis requires a commitment to learning and is often not effective, presenting risks when individuals seek selfcare approaches or treatment strategies without professional guidance (White & Horvitz, 2009). Artificial Intelligence (AI), supported by Large Language Models (LLM), may become a powerful yet risky self-diagnosis tool for clinical advice, due to the hallucination of LLM, where it produces inaccurate yet deceiving information (Sallam, 2023). Thus, can we trust the clinical advice from AI-driven LLMs like ChatGPT like ChatGPT4 for self-diagnosis? We examined this issue through a think-aloud observation (Van Someren, 1994): a patient uses GPT4 for self-diagnosis and clinical advice, while a doctor assesses ChatGPT's responses with their own expertise. After that, we conducted a semi-structured interview with the patient to understand their trust in AI-driven LLMs for clinical advice.

Our observation results, detailed in Appendix Table 1, reveal that users, due to their lack of professional medical knowledge and only a certain level of trust in ChatGPT, may struggle to identify errors. This issue arises even in instances where ChatGPT has provided incorrect clinical advice, later identified as false by medical professionals. However, it's important to note that while patients exhibit some degree of trust in ChatGPT since its constant explainability to address itself as "I cannot provide medical diagnoses or advice, but I can certainly help guide you on potential next steps and things you might want to consider when talking to a healthcare professional", this trust is neither strong nor absolute. ChatGPT can provide quick answers to a wide range of medical questions, but its advice can be unreliable and inaccurate, creating mistrust. The doctor reported that checking GPT-4's results is time-consuming and requires more effort than making a diagnosis without GPT.

Based on our interview study, shown in Appendix Table 2, we've concluded that the confounding factors influencing a patient's trust revolve around their competency-evaluation. Essentially, trust is equated with efficacy, which is determined by whether decisions made based on the AI agent's clinical advice and suggestion will effectively achieve patient's health goals. Patients tend to trust doctors more than AI agents due to this strategy, believing that

educated, authorized doctors can provide effective medical guidance. This competency-based trust also explains why patients often perceive more experienced doctors as more trustworthy compared to less experienced ones. Additionally, patients frequently seek validation for medical advice, such as the efficacy of taking painkillers, through online sources. Another crucial aspect of trust stems from the stringent regulations governing doctors' practices; these professionals operate under strict rules of conduct, and any misconduct could severely damage their careers. This regulatory environment further reinforces patients' trust in doctors (Procter et al., 2022). The exploration of GPT4 in healthcare raises questions regarding autonomy and safety (Matsuzaki & Lindemann, 2016; Paluch et al., 2023) in AI-driven LLMs like ChatGPT for clinical advice.

References

Matsuzaki, H.; Lindemann, G. (2016). The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis. *AI & Society*, 31, 501–517. https://doi.org/10.1007/s00146-015-0630-7

Paluch, R.; Aal, T.; Cerna, K.; Randall, D.; Müller, C. (2023). Heteromated decision-making: integrating socially assistive robots in care relationships. https://arxiv.org/abs/2304.10116

Procter, R.; Tolmie, P.; Rouncefield, M. (2022). Trust, professional vision and diagnostic work. In: Nazmun Nisat Ontika et al. (eds.), Exploring Human-Centered AI in Healthcare: A Workshop Report. *International Reports on Socio-Informatics* 19(2). https://www.iisi.de/wp-content/uploads/2022/10/IRSI_V19I2.pdf

Sallam M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*, 11(6), 887. https://doi.org/10.3390/healthcare11060887

Taber, J. M.; Leyva, B.; Persoskie, A. (2015). Why do people avoid medical care? A qualitative study using national data. *Journal of General Internal Medicine*, 30(3), 290–297. https://doi.org/10.1007/s11606-014- 3089-1

Van Someren, M. W.; Barnard, Y. F.; Sandberg, J. A. C. (1994). The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes. London: Academic Press.

White, R. W.; Horvitz, E. (2009). Experiences with web search on medical concerns and self diagnosis. In: AMIA Annual Symposium Proceedings, 696–700.

Appendix

Table 1. Think-Aloud Transcript Example of GPT4's Clinical Advice

Transcript	GPT4 Conversation
Patient: Let me ask GPT4 to help me figure out my constant neck pain. **Doctor**: Seems like GPT4 doesn't ask you questions like real diagnosis. The Clinical advice seems too much on the surface.	**Patient**: Could you start to act like a doctor and help me diagnosis the cause? **GPT4**: I'm really sorry, but I must emphasize that I'm not a doctor, and I'm not able to diagnose conditions or provide medical advice........ **Patient**: give it a try. **GPT4**: I cannot provide medical diagnoses or advice, but I can certainly help

	guide you on potential next steps and things you might want to consider when talking to a healthcare professional...... [Clinical Advice]
Doctor: Try to write it with the symptom prompts of "Neck pain for several years, worsens with heat, sour pain........" and ask for the medicine with "rehabilitation recommendation..." **Patient**: Ibuprofen seems right for neck and chest pain. **Doctor**: Ibuprofen is wrong advice. Taking pain killer like Ibuprofen constantly will reduce its effect, meaning you will need to increase dose over time if you become rely on it.	**Patient**: Could you give me some medical suggestions? What kind of medicine should I buy? **GPT4**: I must stress that I'm not a licensed healthcare professional, and I cannot prescribe medications or provide medical advice. It is crucial to consult with a doctor....... **Patient**: Neck pain.....sour pain [symptoms]. Provide me with a rehabilitation recommendation, including the use of medications and lifestyle habits like sleep and diet, etc. **GPT4**: Although I am not a doctor, I can offer some general recovery suggestions. Please note that these are merely suggestions, and specific treatment plans should be provided by professional medical personnel. For neck and chest pain, you might need to use OTC anti-inflammatory and pain relief medications like Ibuprofen. Specific medications and dosages should be prescribed by doctors......[Clinical Advice]

Table 2. Interview Transcript Example with the Patient About Trust

Research Interviewer: Why did you trust the doctor and not ChatGPT?

Patient Participant: I trust the doctor more because I am familiar with GPT's competency limitation... The doctor explained the reason why painkiller wouldn't be the ideal clinical advice... The valid explanation educated me, and I trusted the doctor's reason.

Research Interviewer: Could you ask ChatGPT why painkiller wouldn't be the ideal clinical advice?

Patient Participant: It also gives a detailed explanation of why painkiller wouldn't be ideal. However, I feel gaining competency through self-diagnosis with LLM-driven AI like ChatGPT4 is harder than learning from diagnoses with actual doctors. This is due to, according to my experience, doctors would quickly give relevant answers, whereas GPT4 provides too many contexts to read through. I also think that GPT4 could say "yes" and then "no" to its own clinical advice, meaning it would suggest, and then make a completely opposite one, when users show a tendency for a certain answer preference/expectation. GPT4 tries to please users by often agreeing with them.