Balancing Empathy and Accountability: Exploring Friction-In-Design For AI-Mediated Doctor-Patient Communication

Evan Selinger[1,*]

[1]Rochester Institute of Technology, One Lomb Memorial Drive, Rochester, NY, United States of America

Abstract

Empathetic communication between doctor and patient is crucial for building trust. Unfortunately, doctors routinely sound robotic and fall short of the empathetic ideal. Given the systemic issues that give rise to this problem, we may want to consider a new approach: adding generative AI to patient portals. Responsible governance will be needed to deploy the technology ethically and effectively, including establishing procedures for holding doctors accountable for integrating AI-generated content into their messages. In this context, direct and indirect friction-in-design strategies are worth exploring.

Keywords

Empathy, Friction-in-Design, Generative AI

1. The Ongoing Struggle of Doctors to Demonstrate Empathy

Empathy is one of the main ingredients for creating trust between doctors and patients. Not only does empathetic communication help patients feel cared for, respected, and empowered, but it also can promote better medical outcomes. That's because trusting patients tend to be medically compliant and more comfortable sharing sensitive information.

Unfortunately, many doctors fall short of conveying empathy when their patients need it. Physicians can be curt, dismissive, and talk over our heads. They can act like our cliched notions of robots despite medical schools offering communications skills training for decades.

The problem remains, despite ongoing efforts to correct it, because doctors have an extremely demanding job that requires clinical focus and speed. The high stakes situation is a pressure cooker—hardly the right environment for putting yourself in someone else's shoes and seeing how someone's test results can have profound implications that ripple across an entire family. Furthermore, watching patients suffer and having to perform a great deal of bureaucratic labor–seemingly endless paperwork after a long day seeing patients–takes a psychological toll. It's only natural for doctors to want to want to protect themselves from burnout and compassion fatigue by shielding themselves from powerful emotions like empathy.

HHAI-WS 2024: Workshops at the Third International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 10—14, 2024, Malmö, Sweden

*Corresponding author.

emsgsh@rit.edu (E. Selinger)

0000-0002-7298-920 (E. Selinger)

2. Innovating Empathic Communication: Help From Generative AI

Emerging applications of artificial intelligence like generative AI offer new avenues for helping doctors communicate more empathetically. To start assessing these possibilities it is necessary to understand what empathy is, at its core, and the extent to which generative AI can simulate aspects of it.

Human empathy has three main components [1].[1] Imagine a doctor entering a waiting room and seeing a patient pulling at their hair while looking down at the floor right before a procedure. At a mere glance, the physician can tell the patient is upset. That is "cognitive empathy," our ability to identify someone else's emotions. When doctors internalize someone else's feelings, like some of a patient's worry, they experience "emotional empathy."[2] Finally, if doctors feel moved to help patients, maybe say something reassuring to provide comfort, they experience "motivational empathy."

Generative AI lacks emotional and motivational empathy. At most, it can only demonstrate limited abilities associated with cognitive empathy. [3] Given these limits, it is essential to acknowledge that the technology cannot care about patients and their families. To believe otherwise is to overestimate the technology–to impute abilities it lacks and, perhaps, assign it moral duties that it neither has nor can meet. Likely, anyone who makes these mistakes has fallen sway to the cognitive bias of anthropomorphism.

Why, then, is generative AI so effective at producing contextually appropriate, empathetic text like, "I can understand why this would worry you"? The answer is clear. Generative AI can mimic empathy linguistically because it excels at detecting patterns in human language. The data used to train tools like ChatGPT includes literature with empathetic characters and news coverage of people experiencing hardship. Generative AI uses this information to predict which reassuring phrases typically appear when people discuss difficult situations. Again, far from demonstrating the ability to connect with others, this display is merely a simulation of care. Nevertheless, the output is enough to help doctors, and they are the agents who can experience all three components of empathy.

Indeed, there are good reasons to believe doctors can use outputs that mimic empathy to communicate more effectively with patients. Both anecdotal reporting and early scholarly

[1] For related arguments about the limits of AI and empathy, see [2] Perry, Anat. "AI Will Never Convey the Essence of Human Empathy." Nature Human Behaviour 7, no. 11 (November 2023): 1808–9. https://doi.org/10.1038/s41562-023-01675-w

[2] The human glance can take in so much information and function as such a vital source of motivation, that some philosophers argue it has ethical dimensions. See [3] Casey, Edward S. The World at a Glance. Bloomington: Indiana University Press, 2007.

[3] For a study of the abilities related to cognitive empathy conducted with the earlier ChatGPT3, see [4] Sorin, Vera, Danna Brin, Yiftach Barash, Eli Konen, Alexander Charney, Girish Nadkarni, and Eyal Klang. "Large Language Models (LLMs) and Empathy – A Systematic Review." medRxiv, August 7, 2023. https://doi.org/10.1101/2023.08.07.23293769.

research support this hypothesis.

3. AI Saves the Day in the ER

Here is a real-life example of ChatGPT helping a doctor. Dr. Josh Tamayo-Saver faced a dilemma in the emergency room. He was treating a 96-year-old woman who had trouble breathing because her lungs were filled with fluid. Her three children, all senior citizens, were panicking. They followed the medical staff around, asking questions and making requests.

Although the pestering was meant to be helpful, it slowed everyone down and made it hard for Dr. Tamayo-Saver to help all the vulnerable patients in his care. The worst part of the delay was the siblings' insistence that Dr. Tamayo administer an IV to their mother. This option was potentially fatal.

Dr. Tamayo-Saver patiently explained his reasons. The siblings didn't back down. Desperate, he turned to ChatGPT. In seconds, the generative AI composed a clear, detailed, and empathetic explanation–one so good at covering the appropriate treatment protocol that the second-guessing stopped. The most astonishing thing is that the AI projected empathy and did not sound robotic. It opened with, "I truly understand how much you care for your mother, and it's natural to feel concerned about her well-being." [5]

4. Online Applications

The greatest potential for generative AI to help doctors convey empathy isn't in face-to-face situations. A better domain is online medical communication systems, like patient portals. Online communication is increasing, and the volume is exacerbating physician burnout.

Here is the type of scenario that I am envisioning. Doctors receive online notes from their patients. They dictate their replies, and an AI attempts to make the replies sound more empathetic. The physician reviews the updated message, edits it if necessary, and sends the response.

5. Mitigating the Risk of Overreliance

Responsibly adopting generative AI in the manner detailed here will require a detailed governance framework. For example, the following issues will need to be carefully addressed: promoting choice for physicians and patients, maintaining transparency, promoting medical accountability, creating and servicing an appropriate generative AI (e.g., effective, privacy-preserving, etc.), and ensuring fair medical billing practices.

To pick one of these dimensions, the only way to deploy generative AI effectively and responsibly is for doctors to be held fully accountable for all their communications; this includes those partially or wholly written by AI. Ironically, if the technology works well much of the time, a problem can arise. Doctors risk falling sway to automation bias and becoming

complacent. Over time, they may be disinclined to diligently review messages and succumb to overlooking poor responses.

Fortunately, proper safeguards can limit this risk. One promising approach is "friction-in-design."[4] This technique intentionally adds elements to a product or service that make it more time-consuming or challenging to use. For example, X (formerly Twitter) used friction-in-design to ask users to pause before sharing articles they didn't have enough time to read. The goal was to reduce the spread of misinformation by slightly slowing people down so they would engage in more deliberate sharing. Another widespread use of the technique is CAPTCHAs–challenges like identifying objects in images before you can access a website. This delay is meant to prevent bad outcomes like bots scraping data for malicious purposes.

Friction-in-design could be applied here in direct and indirect ways. Direct approaches to friction-by-design options use code to limit the speed at which physicians can reply to patients. By contrast, indirect ones are reminders designed to motivate doctors to spend additional time reviewing correspondence without providing enforcement mechanisms. To further clarify these ideas, let us consider some of the possible direct and indirect ways of designing friction.

6. Friction-In-Design: Direct Options

One direct option is to require *mandatory physician review*. Before sending any AI-mediated messages, doctors should be prompted to carefully review the content. Prompts could be phrased in different ways, and they could include reminders of the professional responsibility to remain accountable and ensure messages accurately reflect their intended communication. Messages would be locked until doctors meet the review requirements. One possible requirement is *timed delays*. The system could enforce a minimum amount of time doctors must spend reviewing messages before they are permitted to be sent.

Yet another direct possibility is to offer *occasional attention checks*. These sporadic prompts could require doctors to answer a brief question or two about the content of messages before they are authorized to send them.

7. Friction-In-Design: Indirect Options

One indirect possibility is for software to *highlight changes* that draw physicians' attention to the specific areas that deserve review. In this scenario, doctors would not need to prove they have examined the changes.

Another indirect possibility is to offer *periodic reminders* of the importance of carefully reviewing AI-generated content and the risks of overreliance. Again, as an indirect option, the mechanism is notice, not enforcement.

[4]For more on friction-in-design, see [6] Brett Frischmann and Susan Benesch "Friction-In-Design Regulation as 21st Century Time, Place, and Manner Restriction" Yale Journal of Law and Technology 25, 376 (2023).

8. Additional Research

Which fiction-in-design approach is best? To answer this question, we need additional research on effectiveness (i.e., the optimal parameters for each approach and how the approaches compare), user experience (i.e., how physicians perceive and judge each option), and untended consequences. This agenda requires interdisciplinary collaboration between experts from medicine, human-computer-interaction, and ethics. It is only by combining insights from these diverse fields that we can create responsible, evidence-based guidelines and reliably foster a deeper understanding of the ethical implications of AI in healthcare communication.

References

[1] C. Montemayor, J. Halpern, A. Fairweather, In principle obstacles for empathic ai: why we can't replace human empathy in healthcare, AI & society 37 (2022) 1353–1359.

[2] A. Perry, Ai will never convey the essence of human empathy, Nature Human Behaviour 7 (2023) 1808–1809.

[3] E. S. Casey, The world at a glance, Indiana University Press Bloomington, 2007.

[4] V. Sorin, D. Brin, Y. Barash, E. Konen, A. Charney, G. Nadkarni, E. Klang, Large language models (llms) and empathy-a systematic review, medRxiv (2023) 2023–08.

[5] J. Tamayo-Sarver, How a doctor uses chat gpt to treat patients, Fast Company (2023). URL: https://www.fastcompany.com/90895618/how-a-doctor-uses-chat-gpt-to-treat-patients, accessed January 8, 2024.

[6] B. Frischmann, S. Benesch, Friction-in-design regulation as 21st century time, place, and manner restriction, Yale JL & Tech. 25 (2023) 376.