Human-Centered Computing
Research Group

FREIE UNIVERSITÄT BERLIN

Protecting privacy when interacting with LLMs in a medical context

Malte Heiser
Freie Universität Berlin
malte.heiser@fu-berlin.de

Jacobe Klein
Freie Universität Berlin
jacobe.klein@fu-berlin.de

Prof. Dr. Claudia Müller-Birn
Freie Universität Berlin
clmb@fu-berlin.de

Context: Challenges of low-acuity patients

- Emotional distress: fear, frustration, pain, and uncertainty about what will happen next.
- Lack of transparency: limited information about waiting times, reasons for delays, and next steps.
- Need for empathy: acknowledgment and reassurance, even with seemingly minor concerns.
- Communication barriers: medical jargon, language differences, and difficulties in describing symptoms.

Vision

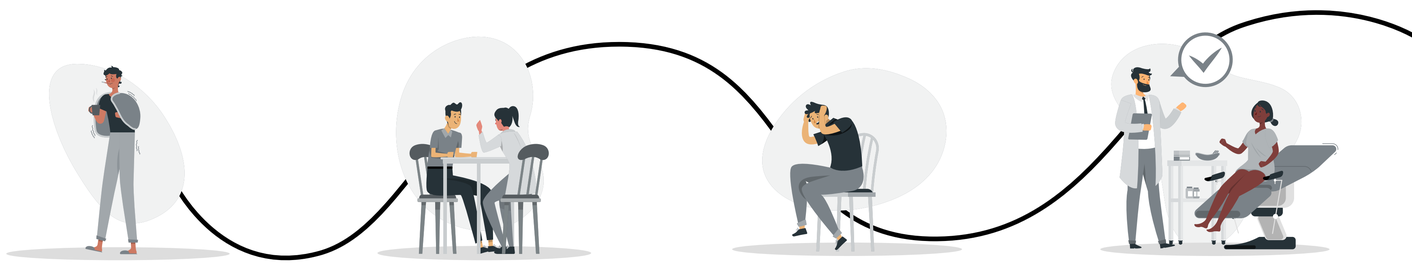
- We envision a short-term LLM-based *ephemeral companion* that supports patients during waiting times in the emergency department (ED).
- Unlike long-term health apps, this companion is designed for brief, situational use: it is present only during the waiting period, offering reassurance, guidance, and preparation for the upcoming consultation.

Core functions

- Reassurance & empathy: reducing anxiety through comforting and supportive interaction.
- Self-reflection & preparation: prompting patients to recall and structure relevant health information.
- Process orientation: explaining ED procedures and clarifying what happens next.
- Information collection: providing clinicians with structured input that can improve efficiency and care quality.

Benefits & Challenges

- Benefits: Preparation and memory aid. Making waiting time meaningful. Orientation and transparency. Emotional support when staff are unavailable.
- Concerns: Limited empathy from LLMs. No diagnoses to not replace clinicians. Usefulness of self-reflection depended on stress levels. **Strong privacy concerns.** [1]



Protecting Patient Privacy

Providing sufficient measures to protect the privacy of LLM-users in the medical domain is paramount to fostering a feeling of security and trust. Only if patients trust the systems they are interacting with, can these systems actually provide benefits.[2]

Relevant Questions

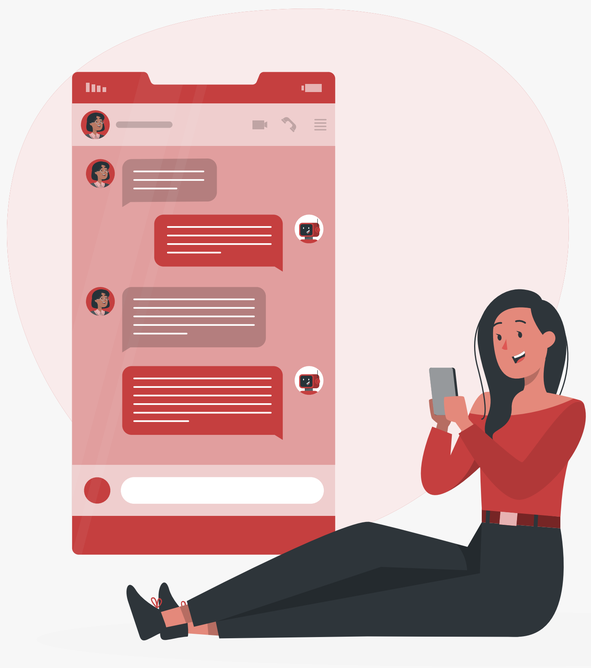
- How can we make sure users are **informed about the LLM-tools** they are using and the associated risks to their privacy?
- How can we encourage the users to **actively protect their data** during interactions with LLMs?
- How can we help users by **limiting transferred personal data** without active user involvement?
- How can we **build safer LLM systems** that can be trained on medical data without the risk of revealing personal information?

Informed Consent

Understanding privacy policies and consent forms is difficult, especially in high-stress situations. Empowering users to make informed decisions could also be done with the help of a chatbot. [3,5]

LLM Safeguards

The best way to protect your data is to not share it at all. Specialized tools could be used to mark potentially sensitive information and empower users to only share the things they are comfortable with.[4,6]



The Privacy-Utility Tradeoff

These same safeguarding measures could also be used to inform the users about the expected tradeoff between the amount of private data shared and the quality of the LLM output.[4,6]

Privacy by Design

For LLMs to improve, they must be fed with as much high-quality data as possible. However, the data must be sufficiently anonymized before training, and the LLM architecture should prohibit inadvertent output of training data.

Acknowledgement

This work is supported by the German Federal Ministry of Research, Technology and Space (grant 0419762107: ILLUMINATION).

Project Website

References

[1] Biswas, A. A., Zulfker, S., Rahman, M., Jani, R. & Anwar, M. Data Privacy and Security Analysis for Mental Health Chatbot Applications. (2025).
[2] Leschanowsky, A., Rech, S., Popp, B. & Bäckström, T. Evaluating privacy, security, and trust perceptions in conversational AI: A systematic review. Computers in Human Behavior 159, 108344 (2024).
[3] Xiao, Z., Li, T. W., Karahalios, K. & Sundaram, H. Inform the Uninformed: Improving Online Informed Consent Reading with an AI-Powered Chatbot. in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems 1–17 (ACM, Hamburg Germany, 2023).
[4] Zhou, J., Xu, E., Wu, Y. & Li, T. Rescriber: Smaller-LLM-Powered User-Led Data Minimization for LLM-Based Chatbots. in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems 1–28 (ACM, Yokohama Japan, 2025).
[5] Freiberger, V., Fleig, A. & Buchmann, E. 'You Don't Need a University Degree to Comprehend Data Protection This Way': LLM-Powered Interactive Privacy Policy Assessment. in Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems 1–12 (ACM, Yokohama Japan, 2025).Jill Palzkill Woelfer, Amy Iverson, David G. Hendry, Batya Friedman, and Brian T. Gill. 2011. Improving the Safety of Homeless Young People with Mobile Phones: Values, Form and Function. Association for Computing Machinery, New York, NY, USA, 1707–1716
[6] Ngong, I. C. et al. Protecting Users From Themselves: Safeguarding Contextual Privacy in Interactions with Conversational Agents. in (2024).
- All illustrations: www.storyset.com.