

Proposal 5

Title:

Empowering Patient Engagement: Leveraging Conversational Explainable AI for Enhanced Healthcare Communication

CONTEXT

Artificial Intelligence (AI) holds transformative potential in healthcare, enhancing diagnostics, treatment plans, and patient outcomes. Despite these advancements, the adoption of AI in healthcare is often hindered by concerns over the transparency and interpretability of AI models. Explainable AI (XAI) addresses these concerns by providing clear explanations of AI decision-making processes. This project focuses on conversational Explainable AI (ConveXAI), utilizing large language models (LLMs) like GPT-3 and ChatGPT to facilitate dialogues between patients and AI systems. By integrating natural language processing with machine learning, ConveXAI aims to demystify AI behaviors, fostering trust and improving patient engagement.

AIM

The objective of this project is to design, develop, and evaluate a prototype of a conversational Explainable AI system that elucidates the behavior of machine learning models in patient care. The research will also explore how Human-Computer Interaction (HCI) principles can enhance end-users' understanding of and trust in these AI systems.

References:

- Leveraging Large Language Models for Patient Engagement.* (2024). arXiv. <https://arxiv.org/html/2406.13659v1>
- Sarkar, S., Gaur, M., Chen, L., Garg, M., Srivastava, B., & Dongaonkar, B. (2023). *Towards Explainable and Safe Conversational Agents for Mental Health: A Survey.* arXiv. <https://arxiv.org/abs/2304.13191>
- Wysocki, O., Davies, J. K., Vigo, M., Armstrong, A. C., Landers, D., Lee, R., & Freitas, A. (2022). *Assessing the Communication Gap Between AI Models and Healthcare Professionals: Explainability, Utility, and Trust in AI-Driven Clinical Decision-Making.* arXiv. <https://arxiv.org/abs/2204.05030>
- Slack, D., Krishna, S., Lakkaraju, H., & Singh, S. (2022). *TalkToModel: Explaining Machine Learning Models with Interactive Natural Language Conversations.* arXiv. <https://arxiv.org/abs/2207.04154>
- Nazary, F., Deldjoo, Y., Di Noia, T., & di Sciascio, E. (2024). *XAI4LLM: Let Machine Learning Models and LLMs Collaborate for Enhanced In-Context Learning in Healthcare.* arXiv. <https://arxiv.org/abs/2405.06270>