AIDING AI-TOOL IMPLEMENTATION IN HEALTHCARE USING RAG

INTRODUCTIONS

AI in healthcare presents significant risks, including bias, potential bodily harm, and the erosion of human autonomy (Mennella et al., 2024). The development and implementation of AI-based prediction models (AIPMs) require strict adherence to ethical and methodological guidelines to ensure responsible use (Nasir et al., 2024). To support this, a six-phase framework has been developed (de Hond et al., 2022), providing structured guidance for AIPM implementation in healthcare. However, this documentation is extensive and often written in highly technical, legal, or academic language, making it challenging for developers, data scientists, and healthcare professionals to apply in practice. To bridge this gap, we are developing a chatbot that assists users in interpreting and applying the guidelines within this framework. The chatbot aims to translate complex information into clear and accessible language (Yang, 2024), facilitating the adoption and correct implementation of AIPMs in healthcare.

Research Questions
How can a chatbot support developers, data scientists, and healthcare professionals in interpreting and applying the framework for implementing AIPMs in healthcare?

Subquestions
- What design principles should be applied to ensure the chatbot is effective and user-friendly for the target audience?
- How can the chatbot-generated information be linked to the original guidelines, providing users with transparent references?
- What factors contribute to end-user trust in the chatbot?

Hypothesis
A chatbot that translates guidelines into comprehensible language improves the accessibility and applicability of the AIPM framework in healthcare. The chatbot will gain more trust from end users when it provides clear and consistent responses, supported by source references.

METHODOLOGY

This study proposes a **multi-method research** approach to develop and evaluate a chatbot that assists developers, data scientists, and healthcare professionals in interpreting and applying the six-phase framework for implementing AI-based prediction models (AIPMs) in healthcare.

The research will begin with a **Systematic Literature Review (SLR)** to identify existing guidelines, challenges, and best practices in AIPM implementation. Following this, **semi-structured expert interviews** will be conducted with AI researchers, developers, and healthcare professionals to gather insights into practical usability needs and limitations.

A **comparative analysis** of existing AI implementation guideline tools and Retrieval Augmented Generation (RAG) solutions will help define best practices and areas for improvement. Based on these insights, we will develop a **prototype**, refining the chatbot's design iterating based on user-feedback.

To evaluate usability and trust, **user-centered qualitative interviews** will be conducted, examining user experience and perceived effectiveness.

ACKNOWLEDGEMENTS

Authors   Yassine El Fahmi, Lucas Maassen

Affiliations   HU Life Sciences & Chemistry
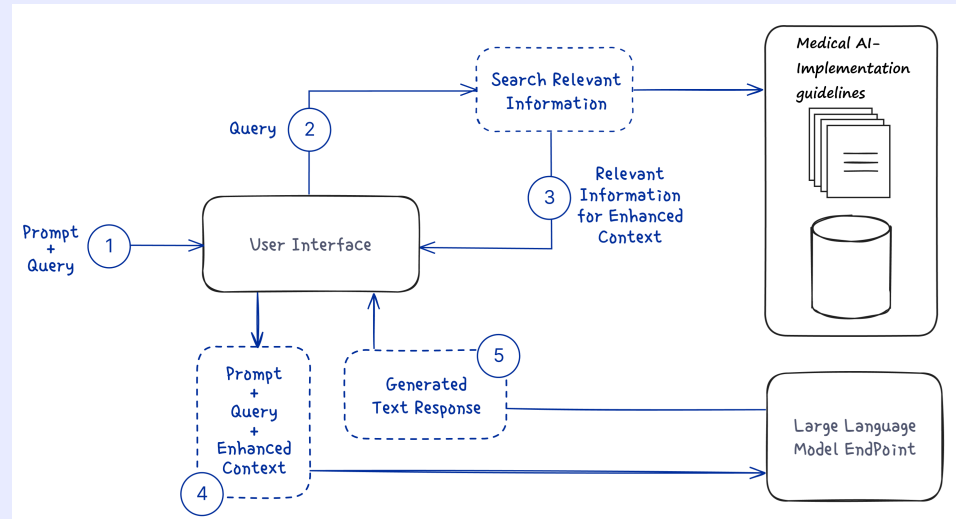University Medical Centrum Utrecht

Figure 1: RAG architecture

LITERATURE REVIEW

Accessing clear and comprehensive guidelines for implementing AI in healthcare remains challenging due to the rapid evolution of AI technologies, which often outpaces the development of standardized protocols and best practices. Existing guidelines are frequently embedded within extensive technical documentation, making them less accessible to healthcare professionals and developers (Nair et al., 2024). Large Language Models (LLMs) have demonstrated significant potential in simplifying complex texts across various domains by rephrasing intricate language into more accessible versions, thereby enhancing comprehension for a broader audience (Yang, 2024). Trust in AI systems is crucial for adoption, especially in healthcare, where fairness, transparency, and explainability are essential (Sheth et al., 2021). Research highlights that human-centered explanations aligned with expert decision-making improve AI trustworthiness (Sheth et al., 2021). However, existing Explainable AI (XAI) methods often lack explicit domain knowledge, making AI decisions difficult to interpret and apply.

REFERENCES

de Hond, A. A. H., Leeuwenberg, A. M., Hooft, L., Kant, I. M. J., Nijman, S. W. J., van Os, H. J. A., Aardoom, J. J., Debray, T. P. A., Schuit, E., van Smeden, M., Reitsma, J. B., Steyerberg, E. W., Chavannes, N. H., & Moons, K. G. M. (2022). Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: A scoping review. Npj Digital Medicine, 5(1), 1–13. https://doi.org/10.1038/s41746-021-00549-7

Mennella, C., Maniscalco, U., Pietro, G. D., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. Heliyon, 10(4). https://doi.org/10.1016/j.heliyon.2024.e26297

Nair, M., Svedberg, P., Larsson, I., & Nygren, J. M. (2024). A comprehensive overview of barriers and strategies for AI implementation in healthcare: Mixed-method design. PLOS ONE, 19(8), e0305949. https://doi.org/10.1371/journal.pone.0305949

Nasir, S., Khan, R. A., & Bai, S. (2024). Ethical Framework for Harnessing the Power of AI in Healthcare and Beyond. IEEE Access, 12, 31014–31035. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3369912

Sheth, A., Gaur, M., Roy, K., & Faldu, K. (2021). Knowledge-Intensive Language Understanding for Explainable AI. IEEE Internet Computing, 25(5), 19–24. IEEE Internet Computing. https://doi.org/10.1109/MIC.2021.3101919

Yang, Z. (2024). Enhancing the Comprehension: Text Simplification Approaches and the Role of Large Language Models. https://scholarshare.temple.edu/handle/20.500.12613/10246