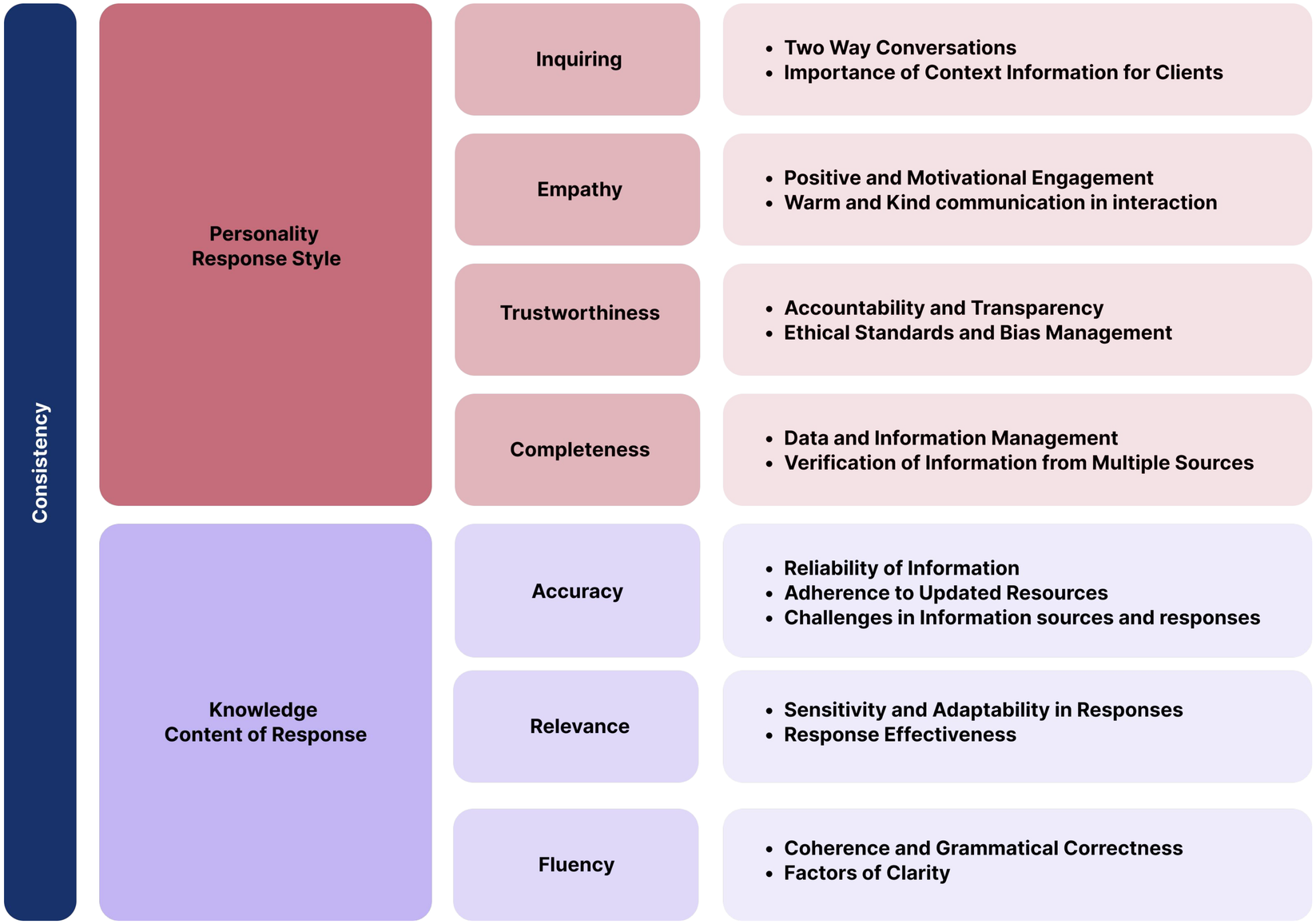
Enabling Domain Expert Evaluation of Emerging AI Technologies in Healthcare Settings



About the study

In this study, we explore the integration of Artificial Intelligence, focusing on Large Language Models (LLMs), within healthcare systems. Our research addresses three main challenges: Identifying the limitations of current evaluation methods, addressing the challenges related to the accuracy and reliability of LLM-generated advice, and proposing improvements to evaluation frameworks to enhance the practical application of these models in healthcare settings.

We employed a chatbot prototype tailored to Dutch healthcare data to test its effectiveness in practical scenarios and refine evaluative measures. Involvement of healthcare professionals in the testing process provided critical, real-world insights into the AI's performance, highlighting the need for dependable and ethically sound AI solutions.

Our findings contribute to the ongoing dialogue on AI in healthcare by proposing a new framework for evaluating AI responses. This framework combines empirical data and theoretical analysis to reduce subjectivity in assessments and standardize the evaluation of AI impacts on healthcare outcomes. Our goal is to develop AI tools that are both responsive and responsible, meeting professional healthcare standards and patient needs.

"Introduction of Dual-Core Evaluation Framework for Conversational AI in Healthcare"

Ujjayan Dhar
Enabling Domain Expert Evaluation of Emerging AI Technologies in Healthcare Settings
Graduation Date : 27.08.2024
MSc Strategic Product Design

Committee Dr. Evangelos Niforatos
Shatha Degachi

Faculty of Industrial Design Engineering

Delft University of Technology