AI Awareness and User Experience in Mental Health Chatbots: A Mixed-Methods Study

Pejman Ebrahimi [1]*, Lynn Miriam Weisker [2], Johannes Schneider [3], Hari Mohan Pandey [4]

[1] Department of Information Systems & Computer Science, University of Liechtenstein, Liechtenstein, ORCID: https://orcid.org/0000-0003-0125-3707, Email: pejman.ebrahimi@uni.li (Corresponding author)

[2] Department of Information Systems & Computer Science, University of Liechtenstein, Liechtenstein, ORCID: https://orcid.org/0009-0005-3621-4670, Email: Lynn.Weisker@uni.li

[3] Department of Information Systems & Computer Science, University of Liechtenstein, Liechtenstein, ORCID: https://orcid.org/0000-0001-7190-9682, Email: johannes.schneider@uni.li

[4] Associate Professor (Sr. Lecturer) of Data Science and Artificial Intelligence, Bournemouth University, United Kingdom, ORCID: https://orcid.org/0000-0002-9128-068X, E-mail: hpandey@bournemouth.ac.uk

Abstract

As artificial intelligence (AI) becomes increasingly integrated into digital mental health care, understanding the nuanced interplay between system transparency, user characteristics, and therapeutic outcomes is critical. This study leverages a robust mixed-methods design with 187 participants to examine how AI transparency, Theory of Change (ToC), advice style, and AI acceptance influence user perceptions of empathy, satisfaction, and treatment outcomes in text-based mental health chatbot interactions. Quantitative analyses revealed no significant main effects of AI transparency or ToC alignment on core outcomes, challenging assumptions about the primacy of system disclosure and user disposition. However, higher pre-existing AI acceptance robustly predicted greater perceived empathy and improved treatment outcomes, particularly in the AI-aware condition. Advanced qualitative analyses, including topic modeling, sentiment analysis, and network-based text mining demonstrated that user discourse was dominated by themes of stress, emotional disclosure, and help-seeking, while chatbot responses consistently reflected supportive and empathetic language. Sentiment distributions and conversational clustering further indicated that the affective and thematic landscape of user-bot interactions remained stable across experimental conditions, with no evidence of enhanced engagement or outcome in the AI-unaware group. These findings underscore the complexity of fostering therapeutic alliance in digital interventions and highlight the need for more personalized, adaptive AI systems that account for individual user expectations and acceptance. The integration of quantitative and qualitative evidence provides actionable insights into the design and deployment of next-generation mental health technologies.

Keywords: *AI Therapy Chatbots, Digital Mental Health, AI Awareness, Theory of Change, User Experience, Sentiment Analysis, Topic Modeling, Therapeutic Alliance*