

The Illusion of Empathy: Why Users Distrust GPT-4 Chatbots for Mental Health Screenings

Tom Bielen

IU International University of Applied Sciences, Belgium

Abstract

Conversational AI holds promise for scalable mental health screening, yet its impact on therapeutic alliance and user perceptions remains underexplored. This preregistered, cross-sectional, randomized mixed-methods experiment ($N = 149$) evaluated the effects of an empathic GPT-4-powered chatbot (Elli) versus a static PHQ-9/GAD-7 form on trust, comfort, empathy, and emotional disclosure. Participants were randomly assigned to either condition. Quantitative outcomes included measures of confidence, comfort, and perceived empathy. Qualitative feedback was analyzed thematically. Primary analyses employed independent t-tests and Mann–Whitney U tests, with Cohen's d effect sizes reported. Exploratory analyses included gender and age interactions, as well as mediation modeling. All analyses were conducted in Python and are openly available on GitHub. Trust in the Elli chatbot was significantly lower than in the static form ($p = .004$, $d = -0.49$). Comfort and empathy ratings showed no significant differences. Dropout analysis revealed no condition-related attrition ($\chi^2 = 0.37$, $p = .54$). Qualitative feedback highlighted discomfort with artificial empathy and a perceived lack of human presence in the chatbot condition. No significant differences emerged for PHQ-9 or GAD-7 severity. Mediation analysis revealed that empathy did not account for the trust gap. Contrary to expectations, the GPT-4 chatbot reduced user trust compared to a static form. These findings suggest emotional authenticity may be more critical than simulated empathy in digital mental health tools.

Keywords: chatbot, gpt-4, digital empathy, trust, mental health, phq-9, gad-7, human–ai interaction