Sympathy for the Devil: Synthetic Empathy Illusion and Safeguards in AI-Mediated Psychotherapeutic Conversations

AI-mediated psychotherapy is emerging not only as a clinical tool but as a phenomenological event: patients may experience chatbot responses as understanding, warmth, and presence. Our empirical findings sharpen this problem: clinicians rated AI-generated psychotherapy conversations more favorably than selected human psychotherapy transcripts, with significantly higher scores for empathy, compassion, and overall preference. Drawing on Husserl's account of intentionality and intersubjective constitution, and Merleau-Ponty's analyses of embodiment and the virtual field of gesture, this paper asks what is given in experience when an artificial system appears empathic, and what this means for clinical practice. I argue that AI can simulate empathic performance while remaining incapable of empathic presence. Clinically, empathy is not merely supportive language but an embodied, affective, and relational orientation in which both the patient and clinician are transformed through encounter. The clinical risk is therefore not only technical error but the illusion of compassion: anthropomorphic trust, dependency, and substitution for human care. I conclude by proposing safeguards—transparency, bounded augmentation rather than replacement, and clinical oversight. The central claim is not that AI cannot comfort, but that comfort is not presence; psychiatry and philosophy must decide together which forms of presence clinical practice may ethically permit.

Word count: 199/200