Transference in artificial intelligence applications

Gamze Zengin Ispir[1]
[1]M.D., University of Health Sciences, Ankara Training and Research Hospital, Alcohol and Drug Treatment Center, Ankara, Turkey
https://orcid.org/0000-0003-3936-6619

Dear Editor,

Woebot: Would you like to share more about what's going on?

User: I feel very strange. There is an indescribable sense of emptiness inside me. When I look at my friends and family, they seem empty too. As a result, we can't connect. There is nothing to hold on to.

Woebot: I understand that you are feeling bad (1).

Artificial intelligence (AI) is on the verge of transforming healthcare significantly by leveraging large datasets to predict potential crises and intervene before they escalate (2). These advancements are shifting treatment processes from reactive to preventive approaches. While AI offers transformative opportunities that reshape mental health practices, it also raises ethical concerns, such as the degree to which human elements should be preserved and issues related to data privacy (2).

In psychiatry, AI applications are redefining the concept of the therapeutic relationship, which has long been a cornerstone of successful mental health treatment (3). Research has shown that patients often reflect their emotional experiences onto AI systems in ways that resemble traditional transference observed between patients and therapists (2, 4). Users of AI therapy chatbots like Replika and Woebot frequently report feelings of emotional trust and attachment despite the absence of human contact. They also tend to share personal and sensitive information as openly as they would with a professional therapist (2). Veterans diagnosed with post-traumatic stress disorder (PTSD) interacting with the virtual therapist Ellie noted that they felt as if they were speaking with a peer and found it easier to share their traumatic experiences due to AI's nonjudgmental nature (4).

AI's nonjudgmental interaction style suggests the possibility of therapeutic collaboration-like dynamics occurring in these interactions. Findings indicate that users may project internalized object relations and self-representations, such as family roles and dynamics, onto AI systems. This new form of transference in AI-assisted therapy should be carefully examined, as it could lead to previously unexperienced psychological outcomes. Although the constant availability of AI therapy bots may offer quick coping strategies for users with depression or anxiety, such accessibility could inadvertently undermine the importance of professional guidance and therapeutic boundaries, potentially fostering dependence on AI (5). Furthermore, for individuals with personality disorders characterized by intense emotions and unstable interpersonal relationships, interactions with AI may reshape emotional responses based on past relational experiences, complicating transference dynamics (6).

For instance, a user who struggles to discuss certain issues due to internalized images of judgmental parents may initially feel relieved when interacting with a virtual therapist who does not judge.

DOI: 10.5505/kpd.2025.60352

Cite this article as: Zengin Ispir G. Transference in artificial intelligence applications. Turkish J Clin Psych 2025; 28:178-180

The arrival date of article: 22.11.2024, **Acceptance date publication:** 28.03.2025

Turkish J Clinical Psychiatry 2025;28:178-180

However, over time, the same user may start belittling or disregarding the virtual therapist due to its perceived lack of judgment, comparing it to their overly critical parents. Another potential transference scenario may arise when the chatbot sends regular check-ins, which could be interpreted as the therapist "remembering" and "caring" for the user. This perception might reinforce rescue fantasies (7). Although transference is usually associated with psychoanalytic psychotherapies, similar dynamics can also emerge in communications with chatbots designed within cognitive behavioral therapy frameworks, such as Woebot and Tess (8).

The remarkable performance of ChatGPT, developed by OpenAI and Microsoft, demonstrates that interactions with large language models are becoming almost indistinguishable from human conversations (1). While some scholars approach these AI capabilities cautiously, others argue that properly guided AI systems could demonstrate reasoning skills and theory of mind-like abilities to perceive and understand emotional states (9, 10). Exploring similarities and differences between AI-assisted applications and human interactions, particularly through the lens of transference, may enhance our understanding of the impacts and future potential of such technologies.

At this stage, examining how transference emerges and takes shape in AI-assisted therapeutic applications is essential. What words, behaviors, or attitudes during user interactions with chatbots trigger transference? Are there specific conversational patterns, questioning styles, or AI communication features that are more likely to induce transference? These are critical questions that warrant further investigation. In psychotherapy, where there is transference, countertransference inevitably follows (7). Can AI systems be trained to recognize users' transference behaviors and adopt a countertransference-like attitude in response? If so, would this "countertransference" correspond to the concept as traditionally understood? More importantly, what contribution could such innovations make to virtual psychotherapeutic practices? Although there is no literature yet addressing countertransference in AI, these questions deserve thorough exploration to better understand AI's role and limitations in therapeutic processes.

Nevertheless, developing AI technologies capable of identifying and managing transference remains challenging. Despite their growing functionality, current AI systems still lack the nuanced ability to fully comprehend and respond to the complex dynamics of psychological processes (2). Furthermore, AI is prone to generating hallucinations—fabricated and incorrect responses independent of user input—which could potentially interfere with transference management (7, 11). For AI to effectively handle transference processes, it must be able to interpret subtle cues and emotional expressions in ways that resemble an experienced therapist's understanding (12). Achieving this requires advances in natural language processing and emotion recognition, as well as the integration of psychological theories into AI frameworks.

Integrating transference into AI-assisted mental health tools also raises ethical concerns, such as the risk of exploiting users' tendencies to seek human connection and reflect emotions for commercial or unethical purposes (2). Moreover, as AI systems become more adept at eliciting and managing transference, boundaries between genuine therapeutic relationships and artificial interactions may blur, potentially leading to harmful attachments. To mitigate these risks, stringent ethical standards should be implemented, prioritizing personal data security, informed consent, transparency, and minimizing dependency on technology.

Ultimately, transference in AI interactions represents a critical intersection between technology and psychiatry in mental health services. The design, programming, and implementation of AI-assisted therapeutic applications should involve close interdisciplinary collaboration among AI developers, psychologists, and psychiatrists (2).

Finally, in the context of discussing transference in AI, the case of IBM's Deep Blue might offer a new perspective. While Deep Blue did not comprehend that it was responding to each move when it defeated chess champion Garry Kasparov in 1997, this did not negate the emotional impact on Kasparov.

Therefore, understanding the emerging dynamics of transference and countertransference with AI applications and their effects on patients and therapists remains crucial.

Future research should focus on developing AI systems capable of adapting to feedback from therapeutic interactions. Such efforts will not only clarify the limitations of AI compared to traditional psychotherapy but also highlight what makes these applications novel and unique.

Correspondence address: M. D., Gamze Zengin İspir, University of Health Sciences, Ankara Training and Research Hospital, Alcohol and Drug Treatment Center, Ankara, Turkey
amatemakademik@gmail.com

REFERENCES

1. Grodniewicz JP, Hohol M. Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence. Front Psychiatry. 2023 Jun 1;14:1190084.

2. Joseph AP, Babu A. Transference and the psychological interplay in AI-enhanced mental healthcare. Front Psychiatry. 2024;15:1460469.

3. Minerva F, Giubilini A. Is AI the future of mental healthcare? Topoi. 2023;42(3):809-817.

4. Sawalha J, Yousefnezhad M, Shah Z, Brown MR, Greenshaw AJ, Greiner R. Detecting presence of PTSD using sentiment analysis from text data. Front Psychiatry. 2022;12:811392.

5. Sadeh-Sharvit S, Camp TD, Horton SE, Hefner JD, Berry JM, Grossman E, Hollon SD. Effects of an Artificial Intelligence Platform for Behavioral Interventions on Depression and Anxiety Symptoms: Randomized Clinical Trial. J Med Internet Res. 2023 Jul 10;25:e46781. doi: 10.2196/46781. PMID: 37428547; PMCID: PMC10366966.

6. Szalai J. The potential use of artificial intelligence in the therapy of borderline personality disorder. J Eval Clin Pract. 2021;27(3):491-496.

7. Özden HC. The Use of Artificial Intelligence in Psychotherapy: Practical and Ethical Aspects. Turk J Psychiatry 2024; 35:336.

8. Prasko J, Ociskova M, Vanek J, Burkauskas J, Slepecky M, Bite I, Krone I, Sollar T, Juskiene A. Managing Transference and Countertransference in Cognitive Behavioral Supervision: Theoretical Framework and Clinical Application. Psychol Res Behav Manag. 2022 Aug 11;15:2129-2155. doi: 10.2147/PRBM.S369294. PMID: 35990755; PMCID: PMC9384966.

9. Floridi L. AI as agency without intelligence: on ChatGPT, large language models, and other generative models. Philos Technol. 2023;36(1):15.

10. Ullman T. Large language models fail on trivial alterations to theory-of-mind tasks. arXiv preprint arXiv:2302.08399. 2023.

11. Özer M. Yapay zekânın varsanıları mı oluyor? Turk Psikiyatri Derg 2024; 35:333-5. https://doi.org/10.5080/u27587

12. Vicci DH. Emotional Intelligence in Artificial Intelligence: A Review and Evaluation Study. Available at SSRN 4818285. 2024.