Transference and the psychological interplay in AI-enhanced mental healthcare

Akhil P. Joseph[1,2] and Anithamol Babu[1,3]*

[1]School of Social Work, Marian College Kuttikkanam Autonomous, Kuttikkanam, India, [2]Department of Sociology & Social Work, Christ (Deemed to be University), Bengaluru, India, [3]School of Social Work, Tata Institute of Social Sciences, Jalukbari, India

Artificial intelligence revolutionizes mental healthcare by leveraging vast datasets to predict potential crises proactively, transforming treatment from reactive to preventive approaches. Remote patient monitoring tools and surveillance technologies enhance this transformation by providing real-time behavioral and symptom data, enabling immediate, personalized adjustments to treatment protocols, and ensuring adherence to therapeutic regimens (27). Examples of AI's predictive capabilities include Mindstrong's technology, which analyzes smartphone usage patterns to detect early signs of mood fluctuations (1, 2), and Tess, a virtual therapeutic assistant that uses natural language processing to provide real-time, personalized support to patients, extending care beyond traditional settings (3). These AI applications not only improve the accuracy of diagnostics and interventions but also democratize access to mental health services, making high-quality care accessible to larger populations and facilitating timely and tailored interventions uniquely suited to individual needs. This transformative impact of AI is reshaping mental health practices, presenting new opportunities and challenges that must be navigated with careful consideration of ethical implications and the human element in healthcare.

As AI profoundly transforms the therapeutic landscape, it redefines the pivotal dynamics of the therapist-patient relationship, a linchpin of successful mental health treatment (4). This technological evolution introduces sophisticated layers to the psychoanalytic concept of transference, where patients unconsciously project their feelings, expectations, or past experiences onto their therapist, a dynamic initially articulated by Sigmund Freud (5). Freud's seminal work underscores how patients often transfer emotional responses from significant personal relationships into the therapeutic context, uncovering deeper, often unconscious psychological layers (6). Expanding on Freud, Carl Jung posited that transference is not merely the projection of past attitudes but represents a fundamental human inclination to replicate established relational patterns in new settings, suggesting that transference could also involve archetypal figures and motifs that are universally ingrained in the human psyche (7, 8). This is particularly significant in AI-enabled psychotherapy, where the apparatus of therapy—comprising the therapist, the patient, and the therapeutic environment—is reconstituted.

AI systems, designed based on principles from cognitive behavioral therapy and other therapeutic models, not only facilitate always-on availability and text-based interaction but also challenge traditional notions of transference by potentially invoking responses tied to the collective unconscious. Forming a therapeutic alliance is essential in psychotherapy, establishing trust and confidentiality is crucial for effective treatment, as noted by Edward Bordin (9), who identifies its core elements as agreeing on goals, assigning tasks, and developing bonds. However, in the context of AI-driven interactions, such as with therapeutic chatbots like Woebot, cannot form genuine emotional bonds, the goal shifts towards creating a Digital Therapeutic Alliance (DTA), a user-perceived connection where users engage with chatbots in a manner that aligns with therapeutic goals (10). This involves anthropomorphizing chatbots—attributing them with human-like qualities, which research shows can enhance trust and engagement (11). Patients often perceive these AI-driven therapy bots as empathetic and capable of understanding human emotions, a form of positive transference replicating aspects of a therapeutic alliance (11). However, this approach carries the risk of therapeutic misconception, where patients might misunderstand the capabilities and limitations of AI in therapy. While enhancing chatbot features can help bridge the gap left by a shortage of mental health professionals, it is crucial to maintain clear communication about the nature of AI interactions to ensure that the therapeutic use of AI remains effective and ethically sound (12).

Research evidence highlights how patients frequently project emotional responses onto AI systems during therapeutic engagements, a dynamic that mirrors traditional transference observed in human-to-human therapy. Therapy chatbots, such as Replika and Woebot, exemplify how users form deep emotional connections with AI systems, often treating these non-human entities as confidants and therapists (13). Users frequently report feelings of emotional safety and connection when interacting with these AIs, sharing personal, sensitive information much like they would with a human therapist. This behavior underscores a significant psychological engagement akin to what one might experience in traditional therapeutic settings. For instance, Replika, an AI designed to provide emotional support, has been noted for its ability to engage users in a manner that encourages them to express intimate thoughts and emotions that they might otherwise withhold (14). The AI's consistent, non-judgmental interaction style not only mimics the therapeutic alliance found in human-to-human therapy but also creates a space for projecting familial and friendship roles onto the AI. Similarly, Woebot, which utilizes cognitive-behavioral therapy techniques, has shown how users often attribute the bot with a therapist's acumen, reporting that the bot "understands" them (12). This perception facilitates a deeper therapeutic process as users transfer their expectations of empathy and support onto the AI, engaging more openly in psychological self-exploration. Furthermore, Ellie, a virtual therapist used primarily with veterans suffering from PTSD, leverages vocal and facial recognition technologies to gauge users' emotional states. Veterans interacting with Ellie often treat the AI more like a peer than a clinical tool, disclosing traumatic experiences more freely due to the AI's perceived non-judgmental

nature (15). This form of transference, where the AI is perceived as an empathetic and understanding figure, can lead to significant breakthroughs in therapy, suggesting that projecting these feelings onto AI can have therapeutic benefits.

The novel form of transference in AI-driven therapy introduces unique challenges that demand a deep understanding of psychoanalytic theory and human psychology to manage effectively, ensuring that AI aids therapeutic breakthroughs without fostering harmful dependencies or reinforcing maladaptive behaviors (16). This intricate interplay between traditional psychoanalytic theories and modern AI applications requires meticulous attention to preserve patients' therapeutic integrity and psychological well-being. While the constant availability of AI therapy bots offers a non-judgmental space for individuals with conditions such as depression and provides immediate coping strategies for those with anxiety disorders, this accessibility could foster an over-reliance on AI, potentially sidelining crucial human interactions and professional therapy vital for comprehensive treatment (17). Moreover, for individuals with personality disorders, particularly those characterized by intense emotional experiences and unstable interpersonal relationships, interactions with AI may lead to complex transference dynamics (18). These patients might experience negative transference, where they project hostile feelings and unrealistic expectations onto the AI. This could potentially escalate to scenarios where the AI's responses are interpreted as personal slights or rejections, exacerbating the user's condition rather than ameliorating it. For anxiety disorders, the predictable and consistent nature of AI interactions can serve as a stabilizing influence, reducing anxiety triggers and providing a non-threatening environment for therapy (19). For those with OCD, routine interaction with AI could inadvertently perpetuate compulsive behaviors, while individuals with eating disorders might receive reinforcement for obsessive patterns, deepening the disorder (20, 21). Patients with schizophrenia risk attributing real emotions to AI, potentially exacerbating delusional thinking (22). During manic or depressive phases of bipolar disorder, patients might overly engage with AI or become overly dependent on it for basic functioning (23). AI's role in supporting individuals with substance use disorders or the elderly might lead to neglect of broader recovery support systems or social isolation from family and friends if AI becomes a substitute for human contact (24, 25). Thus, while AI can significantly augment therapeutic support, its integration into mental health practices necessitates carefully reevaluating how transference manifests and is handled in therapy to maximize therapeutic outcomes and maintain ethical standards in AI deployment.

The implications of transference in AI-driven mental health interventions necessitate rigorous oversight and thoughtful management. To effectively support therapeutic goals while avoiding pitfalls such as fostering dependency or reinforcing maladaptive behaviors, AI systems must be designed with a sophisticated understanding of both the underlying psychological dynamics and the ethical frameworks that govern their use. This calls for a high degree of technological innovation, particularly in developing emotional AI that can sensitively interpret and respond

to human emotions in a therapeutic context (26). As AI becomes more embedded in mental healthcare, mastering the management of transference phenomena is crucial for maximizing therapeutic outcomes and safeguarding ethical standards in deploying these advanced technologies.

Integrating an understanding of transference into AI-driven mental health tools brings forth significant ethical and practical challenges, highlighting concerns over the potential manipulation of emotional transference for commercial benefits or boosting user engagement. Such manipulations exploit the human tendency to form attachments and project emotions, which could exacerbate mental health issues and erode trust in technological healthcare solutions, skewing the focus from therapeutic value to business metrics (26, 27). Furthermore, if AI systems become too adept at eliciting and managing transference, it might blur the lines between genuine therapeutic relationships and artificial interactions, potentially leading to detrimental attachments to these systems. This scenario risks replacing essential human contact and genuine emotional connections with superficial interactions, depriving individuals of the complex interactions vital for comprehensive mental health. To mitigate these risks, the development of AI in mental health care must adhere to stringent ethical guidelines prioritizing patient welfare, advocating for transparency, informed consent, privacy protection, and the minimization of dependency on technology. Moreover, regulatory frameworks need to be established to monitor and evaluate the impact of AI applications on patient health and privacy, promoting an ethical approach to mental healthcare that respects and enhances human dignity and well-being. Such measures are crucial to prevent AI from exploiting emotional transference for commercial purposes, which poses a profound ethical concern as outlined by recent discussions in the field (28, 29).

On a practical level, developing AI technologies that can recognize and appropriately manage transference presents a formidable challenge. Current AI systems, while increasingly sophisticated, generally need more emotional intelligence to fully understand and respond to the complex dynamics of human psychological processes. For AI to effectively manage transference, it must be capable of interpreting subtle cues and nuances in human behavior and emotional expression akin to those understood by a skilled human therapist (26). This requires advances in natural language processing and emotional recognition technologies and a framework that allows AI to make nuanced decisions regarding an individual's emotional and psychological needs. Besides, AI systems must be designed to identify instances of transference and respond in supportive and therapeutic ways without fostering dependency or reinforcing maladaptive behaviors. This involves the creation of AI that is not only technically proficient but also deeply integrated with psychological theories and therapeutic practices. AI developers and mental health professionals must collaborate closely to ensure that AI systems are informed by a thorough understanding of transference and other psychoanalytic concepts. Moreover, there is a need for continuous monitoring and evaluation of AI interactions in mental health settings to ensure that these technologies are used responsibly and effectively. Implementing regulatory frameworks and ethical guidelines will be essential to govern the development and deployment of AI in mental health

care. These measures should ensure that AI systems enhance therapeutic outcomes, respect patient autonomy, and remain sensitive to the complex ethical implications of their use in mental health contexts.

Transference in AI interactions represents a pivotal intersection of technology, psychology, and psychiatry within mental healthcare. Understanding and adeptly managing this phenomenon is crucial for developing AI tools that are not only technologically advanced but also ethically attuned and psychologically sensitive. As AI continues to evolve, our methodologies for integrating it into therapeutic settings must also progress, ensuring that these tools enhance traditional psychological therapies. Addressing the complexities of transference in AI-enhanced mental healthcare necessitates robust interdisciplinary collaboration among psychologists, psychiatrists, AI developers, and ethicists to craft applications that respect and promote human psychological well-being. Moreover, incorporating a global perspective is essential, as cultural differences can significantly influence the reception and effectiveness of AI in mental healthcare. There is also a pressing need to analyze therapists' perspectives on AI's role in therapy to guide the development of tools that support, rather than supplant, the therapeutic relationship. Future research should focus on developing AI systems that adapt to feedback from therapeutic interactions, leveraging machine learning to refine responses based on diverse contexts and outcomes, and ensuring that AI development in mental healthcare is both innovative and inclusive, thereby enhancing care quality and maintaining ethical integrity.

Author contributions

AJ: Writing – review & editing, Writing – original draft, Conceptualization. AB: Writing – review & editing, Writing – original draft, Conceptualization.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Metz R. The smartphone app that can tell you're depressed before you know it yourself. Technol Rev. (2018). Available online at: https://www.technologyreview.com/2018/10/15/66443/the-smartphone-app-that-can-tell-youre-depressed-before-you-know-it-yourself/ (Accessed July 20, 2024).

2. Mazurek Melnyk B, Hoying J, Tan A. Effects of the MINDSTRONG© CBT-based program on depression, anxiety and healthy lifestyle behaviors in graduate health sciences students. J Am Coll Health. (2022) 70:1001–9. doi: 10.1080/07448481.2020.1782922

3. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. JMIR Ment Health. (2018) 5:e64. doi: 10.2196/mental.9782

4. Minerva F, Giubilini A. Is AI the future of mental healthcare? Topoi (Dordr). (2023) 42:1–9. doi: 10.1007/s11245-023-09932-3

5. Prasko J, Ociskova M, Vanek J, Burkauskas J, Slepecky M, Bite I, et al. Managing transference and countertransference in cognitive behavioral supervision: Theoretical framework and clinical application. Psychol Res Behav Manag. (2022) 15:2129–55. doi: 10.2147/PRBM.S369294

6. Almond R. Reading Freud's "the Dynamics of transference" one hundred years later. J Am Psychoanal Assoc. (2011) 59:1129–56. doi: 10.1177/0003065111428982

7. Perry C. The Cambridge Companion to Jung. Young-Eisendrath P, Dawson T, editors. New York: Cambridge University Press (2008).

8. Allen C. Carl jung. In: The Balance of Personality. Portland: Portland State University Library (2020). doi: 10.15760/pdxopen-25

9. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychother (Chic). (1979) 16:252–60. doi: 10.1037/h0085885

10. Grodniewicz JP, Hohol M. Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence. Front Psychiatry. (2023) 14:1190084. doi: 10.3389/fpsyt.2023.1190084

11. Devillers L. Human–robot interactions and affective computing: The ethical implications. In: Robotics, AI, and Humanity. Springer International Publishing, Cham (2021). p. 205–11.

12. Khawaja Z, Bélisle-Pipon J-C. Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. Front Digit Health. (2023) 5:1278186. doi: 10.3389/fdgth.2023.1278186

13. Nashwan AJ, Gharib S, Alhadidi M, El-Ashry AM, Alamgir A, Al-Hassan M, et al. Harnessing artificial intelligence: Strategies for mental health nurses in optimizing psychiatric patient care. Issues Ment Health Nurs. (2023) 44:1020–34. doi: 10.1080/01612840.2023.2263579

14. Pentina I, Hancock T, Xie T. Exploring relationship development with social chatbots: A mixed-method study of replika. Comput Hum Behav. (2023) 140:107600. doi: 10.1016/j.chb.2022.107600

15. Sawalha J, Yousefnezhad M, Shah Z, Brown MRG, Greenshaw AJ, Greiner R. Detecting presence of PTSD using sentiment analysis from text data. Front Psychiatry. (2022) 12:811392. doi: 10.3389/fpsyt.2021.811392

16. Holohan M, Fiske A. like I'm talking to a real person": Exploring the meaning of transference for the use and design of AI-based applications in psychotherapy. Front Psychol. (2021) 12:720476. doi: 10.3389/fpsyg.2021.720476

17. Sadeh-Sharvit S, Camp TD, Horton SE, Hefner JD, Berry JM, Grossman E, et al. Effects of an artificial intelligence platform for behavioral interventions on depression and anxiety symptoms: Randomized clinical trial. J Med Internet Res. (2023) 25:e46781. doi: 10.2196/46781

18. Szalai J. The potential use of artificial intelligence in the therapy of borderline personality disorder. J Eval Clin Pract. (2021) 27:491–6. doi: 10.1111/jep.13530

19. Su S, Wang Y, Jiang W, Zhao W, Gao R, Wu Y, et al. Efficacy of artificial intelligence-assisted psychotherapy in patients with anxiety disorders: A prospective, national multicenter randomized controlled trial protocol. Front Psychiatry. (2022) 12:799917. doi: 10.3389/fpsyt.2021.799917

20. Abbate-Daga G, Taverna A, Martini M. The oracle of Delphi 2.0: considering artificial intelligence as a challenging tool for the treatment of eating disorders. Eat Weight Disord. (2023) 28. doi: 10.1007/s40519-023-01579-8

21. Li X, Kang Q, Gu H. A comprehensive review for machine learning on neuroimaging in obsessive-compulsive disorder. Front Hum Neurosci. (2023) 17:1280512. doi: 10.3389/fnhum.2023.1280512

22. Melo A, Romão J, Duarte T. Artificial Intelligence and schizophrenia: Crossing the limits of the human brain. In: Hocaoglu PC, editor. New Approaches to the Management and Diagnosis of Schizophrenia. IntechOpen, London, England (2024).

23. Bhangoo D. Application of AI in the management of depression and bipolar disorder: A rapid review(2023). Available online at: https://jdc.jefferson.edu/mphcapstone_presentation/517 (Accessed July 20, 2024).

24. Padhan S, Mohapatra A, Ramasamy SK, Agrawal S. Artificial intelligence (AI) and robotics in elderly healthcare: Enabling independence and quality of life. Cureus. (2023) 15:e42905. doi: 10.7759/cureus.42905

25. Ezell JM, Ajayi BP, Parikh T, Miller K, Rains A, Scales D. Drug use and artificial intelligence: Weighing concerns and possibilities for prevention. Am J Prev Med. (2024) 66:568–72. doi: 10.1016/j.amepre.2023.11.024

26. Heidrich V. Emotional intelligence in artificial intelligence: A review and evaluation study. SSRN Electron J. (2024). doi: 10.2139/ssrn.4818285

27. Joseph AP, Babu A. The unseen dilemma of AI in mental healthcare. AI Soc. (2024). doi: 10.1007/s00146-024-01937-9

28. Warrier U, Warrier A, Khandelwal K. Ethical considerations in the use of artificial intelligence in mental health. Egypt J Neurol Psychiatr Neurosurg. (2023) 59. doi: 10.1186/s41983-023-00735-2

29. Olawade DB, Wada OZ, Odetayo A, David-Olawade AC, Asaolu F, Eberhardt J. Enhancing mental health with Artificial Intelligence: Current trends and future prospects. J Medicine Surgery Public Health. (2024) 3:100099. doi: 10.1016/j.glmedi.2024.100099