Psychotherapist Bots: Transference and Countertransference Issues

Fatemeh AMIRKHANI [a] ,Zahra NOROUZI [b], and Saeedeh BABAII [c,1]
[a] *Ph.D. Student in Psychology, University of Allameh Tabataba'i, Tehran, Iran*
[b] *Ph.D. Student in Clinical Psychology, University of Shahed, Tehran, Iran*
[c] *Postdoc Fellow, International Center for Ethics in the Sciences and Humanities (IZEW), University of Tuebingen, Germany*

International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL

Keywords: psychotherapist bots, transference countertransference, chatbots

Abstract

There is a rapid advancement in the development of psychotherapist bots that are based on artificial intelligence. Chatbots and robots may facilitate treatment by reducing barriers and increasing accessibility. Researchers have shown that psychological bots play an effective role similar to traditional face-to-face psychotherapy in reducing depression and anxiety symptoms. Due to the rapid advancement of psychotherapy technology, therapeutic chatbots are likely to become widely used in the near future. In this context, it is essential to consider both the ethical and clinical aspects of bots and chatbots as mental healthcare improvement assistants. The first part of this abstract outlines the concept of transference and countertransference in human-psychotherapist bot interactions. In this novel form of therapy, topics like transference and countertransference need to be discussed, as well as concepts such as empathy, acceptance, judgment, and safety in therapeutic relationships. We attempt to draw attention to the need to revisit clinical and ethical issues related to the interactions between humans and psychotherapist bots.

Introduction

There is a rapid advancement in the development of psychotherapist bots that are based on artificial intelligence. Chatbots and bots (collectively known as bots) may facilitate treatment by reducing barriers and increasing access to care. During the COVID-19 pandemic, psychotherapist bots provided a critical alternative to in-person healthcare (Miner et al., 2020). In numerous studies, chatbots like Tess, Wysa, Sara, or Woebot have been shown to be as effective as classic face-to-face psychotherapy methods, for instance, cognitive-behavioral techniques (Fiske et al., 2019). Researchers found that psychotherapist bots reduced depression and anxiety symptoms (Bendig et al., 2019). Therapeutic bots will likely become much more widely used in clinical applications in the near future, given how rapidly technologies have evolved in psychotherapy (Holohan & Fiske, 2021). While bots as a tool for psychotherapy are still at an early stage of development, there are likely to be a number of important questions to be answered.

---

[1] corresponding author, E-mail: saeedeh.babaii@izew.uni-tuebingen.de, Saeedeh Babaii's work was funded by the Baden-Württemberg Stiftung (`Verantwortliche Künstliche Intelligenz') as part of the AITE project.

In human-psychotherapist bot interactions, one critical topic is “transference and countertransference” which are characteristics of therapeutic relationships (Fiske et al., 2019). Traditionally, transference refers to the unconscious feelings of a client toward a significant other, such as their therapist. The transference of feelings/reactions offers valuable insight into a client's inner world that will be discussed further in the paper. Countertransference describes the feelings therapists may have for their clients when working with them (Jenks & Oka, 2020). Both transference and countertransference are related to a human's ability to perceive and recognize the outside world. When someone perceives something for the first time, s/he needs to compare the present perception with what s/he has experienced previously to recognize it and have a better understanding of the world.

As AI systems become more emotionally engaging, these concepts become increasingly relevant. AI systems are designed and trained to elicit, recognize, and sometimes simulate human emotions (Weber-Guskar, 2021) and interact emotionally and socially with users in settings such as therapeutic relationships. In this abstract, we first outline the concept of transference and countertransference in human/psychotherapist bot interaction. Then we discuss some ethical and psychological issues related to transference and countertransference in this new form of therapy.

Transference and Countertransference with Psychotherapist Bots

Based on the studies on bots, we can see that some users develop human-like connections with their bots (Holohan & Fiske, 2021). Bots become therapeutic only through their interaction with users, who in this case are clients looking for psychological help (Haraway, 2008). It is therefore readily apparent that the apparatus has changed when the therapist is no longer a human but a bot. Since transference is crucial to the psychotherapeutic apparatus, we must ask how AI-driven innovations could be designed to account for and even foster opportunities for transference in a useful way. In other words, it will be necessary to conceive of transference not as an unanticipated byproduct of AI-directed psychotherapy, but to actively consider it in the design process and limit it to an ethically acceptable extent (Holohan & Fiske, 2021).

The transference in the newly created setting can occur in two relationships:

1- designer/bot: transference in design

> The bot was constructed and designed in such a way that traces of transference can be found. The bot's characteristics can be influenced by the designer's relationship with his parents as a child (Dietrich, Fodor, Zucker & Bruckner, 2009).

2- client/bot

- Transference in therapeutic communication

Humans experience negative and positive transferences such as love, hate, admiration, fear, closeness, and distance in relation to the bot therapist, as they do in relation to other humans. We know that the transference process can be formed through the avatar or wording and the way the bot talks to the person (Holohan & Fiske, 2021). There are still questions such as whether the transference experience in relation to the bot therapist is exactly the same as the experience with a human therapist. How might the client relate transferentially to the bot (through what words, behaviors, demeanor, etc.)? In approximating human therapist responses, are there specific speech patterns, forms of

questioning, or other features of AI communication that might give rise to specific forms of transference in the therapeutic encounter?

- Transference in therapeutic setting

Traditionally, the client comes to the therapist's office for treatment. As for the bot, the client can talk to the bot while lying on the bed in pajamas and without make-up or while sitting in the classroom (Butterworth, 2020). These new dynamics can shape different modes of transference experiences in relation to the bot.

As of yet, no research has been conducted about countertransference between humans and bot therapists. There is a possibility that the bot will show the same transference to everyone due to data bias (Szalai, 2021). In addition, countertransference from the bot can make the psychotherapist bot more in control, increase manipulation possibilities, and end the treatment process.

Transference and Countertransference Issues with Psychotherapist Bots

Therapeutic alliance

Therapeutic alliances can be fostered by transference, especially during the early stages. Positive transference can make it possible for the client to face difficult subjects, by helping them feel supported and understood (Holohan & Fiske, 2021). Further, by transferring emotions between the client and the bot, the client can experience a sense of belonging (Dietrich, Fodor, Zucker & Bruckner, 2009), which may lead to ethical dilemmas.

Empathy

Empathy is required in addressing transference and countertransference in order to respond authentically and compassionately to the client's suffering. The person can experience true empathy from the therapist once the relationship integrates and transcends past relational representations. The therapist uses current countertransference reactions to understand the client's inner life (Norton, 2011). As an argument against integrating bots into mental health practices, it has been argued that they cannot empathize the same way as humans (Montemayor et al., 2021). During in-person therapy, the brain focuses partly on the words spoken, but it also takes into account dozens of non-verbal signals and can lay the groundwork for intimacy (Sayers, 2021). Psychotherapist bots will be able to form therapeutic alliances with their users as they display empathy, curiosity, and understanding. As technology incorporating empathy and emotions continues to advance, it brings up a range of ethical concerns (Fosch Villaronga, 2019).

Judgment and acceptance

Clients who use psychotherapeutic bots may experience a sense of relief from not feeling judged, as they are aware that they are interacting with a non-judgmental bot. On the one hand, this creates a sense of safety for them, thereby making it easier to discuss challenging subjects. On the other hand, the client might at the same time contrast this absence of judgment with the overly judgmental attitude of their mother. The client might ultimately fail to take their bot therapist seriously, or even treat it with disdain because, through their transference, they ascribe a lack of authority to the bot, even though interacting with it makes them feel safe and cared for (Holohan & Fiske, 2021).

Likewise, if a client experiences paranoid thoughts, they may perceive acceptance from a bot as a threat. Even when disclosure is made that the system is “just a machine”, some clients may believe that the machine is “alive” or that there is a person, or some other malevolent force behind the simulation (Luxton, 2014).

Safety

In a therapeutic relationship, the client may inadvertently disclose their problems to the therapist (transference). It is essential that the therapist allows the client to use the therapist as a symbol and as a container for conflicted and unresolved emotions. The transference of emotional issues must be protected and analyzed. Therefore, safeguarding confidentiality, privacy, sensitive information, and taboo subjects are pivotal aspects of therapy (Gordon, 1993). Presently, there are no specific confidentiality or privacy standards and regulations in place for chatbots and other mental healthcare bots (Stiefel, 2018).

Furthermore, AI-powered algorithms in psychotherapist bots may inadvertently introduce data-driven biases based on gender, race, type of the mental issue, etc., resulting in unintended harm or exclusion of certain individuals (Fiske et al., 2019). Clients can be exposed to these biases in a variety of ways through countertransference. Psychotherapist bots should promote rules of neutrality and transparency so that their own biases or deficiencies will not interfere with treatment.

Conclusion

Transference and countertransference are significant concepts in the psychotherapeutic relationship. As bots continue to advance, the concept of transference and countertransference is also applicable to the interaction between humans and bots. As discussed in this article, transference plays a critical role in designing, constructing, and operating bots, therapeutic interactions, and therapeutic settings. Various transference-related concerns were addressed, such as the therapeutic alliance, empathy, safety, and judgment and acceptance, highlighting the importance of considering them in the future design of psychotherapist bots

References

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The next generation: Chatbots in clinical psychology and psychotherapy to Foster Mental Health – a scoping review. Verhaltenstherapie, 1–13.

Butterworth, J. (2020). Online Therapy: A New Dimension to Transference and Countertransference. Journal of the New Zealand College of clinical psychologist, 30(1), 7-8.

Dietrich, D., Fodor, G., Zucker, G., & Bruckner, D. (Eds.). (2009). Simulating the mind: A technical neuropsychoanalytical approach. Vienna: Springer Vienna.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your bot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. Journal of Medical Internet Research, 21(5).

Fosch Villaronga, E. (2019). “I love you,” said the bot: Boundaries of the use of emotions in human-bot interactions. Human–Computer Interaction Series, 93–110.

Gordon, R. M. (1993). Ethics based on protection of the transference. Issues in Psychoanalytic Psychology, 15(2), 95–105.

Haraway, D. J. (2008). When Species Meet. Minneapolis: University of Minnesota Press.

Holohan, M., & Fiske, A. (2021). “like I’m talking to a real person”: Exploring the meaning of transference for the use and design of AI-based applications in psychotherapy. Frontiers in Psychology, 12.

Jenks, D. B., & Oka, M. (2020). Breaking hearts: Ethically handling transference and countertransference in therapy. The American Journal of Family Therapy, 49(5), 443–460.

Luxton, D. D. (2014). Recommendations for the ethical use and design of artificial intelligent care providers. Artificial intelligence in medicine, 62(1), 1-10.

Miner, A. S., Laranjo, L., & Kocaballi, A. B. (2020). Chatbots in the fight against the COVID-19 pandemic. Npj Digital Medicine, 3(1).

Montemayor, C., Halpern, J., & Fairweather, A. (2021). In principle obstacles for empathic AI: why we can’t replace human empathy in healthcare. AI &Amp; SOCIETY.

Norton, C. L. (2011). Developing empathy: A case study exploring transference and countertransference with adolescent females who self-injure. Journal of Social Work Practice, 25(1), 95–107.

Prasko, J., Diveky, T., Grambal, A., Kamaradova, D., Mozny, P., Sigmundova, Z., Slepecky, M., & Vyskocilova, J. (2010). Transference and countertransference in cognitive behavioral therapy. Biomedical Papers, 154(3), 189–197.

Sayers, J. (2021). Online psychotherapy: Transference and countertransference issues. British Journal of Psychotherapy, 37(2), 223–233.

Stiefel, S. (2018). 'The Chatbot Will See You Now': Mental Health Confidentiality Concerns in Software Therapy. SSRN Electronic Journal.

Szalai, J. (2021). The potential use of artificial intelligence in the therapy of borderline personality disorder. Journal of Evaluation in Clinical Practice, 27(3), 491-496.

Weber-Guskar, E. (2021). How to feel about emotionalized artificial intelligence? when bot pets, holograms, and chatbots become affective partners. Ethics and Information Technology, 23(4), 601–610.