Turkish Journal of Psychiatry 2024;35(4):336–339

Türk Psikiyatri Dergisi

Turkish Journal of Psychiatry

Letter to the Editor

The Use of Artificial Intelligence in Psychotherapy: Practical and Ethical Aspects

Dear Editor,

Artificial intelligence (AI) has the potential to create radical changes in the field of psychiatry. Examples of the positive effects of AI on psychiatry include innovations in the diagnostic process, development of new treatment methods, personalized treatment approaches, and demonstration of treatment effectiveness with objective data (Ayhan 2023). However, in addition to all these advantages, there are also concerns about the use of AI in psychiatry. Ethical and confidentiality concerns regarding the use of AI, uncertainties regarding the transparency of decision processes, misleading results, and the risk of shaping treatment according to these results are the main concerns expressed (Ayhan 2023, Gürcan et al. 2024).

AI has the potential to influence and change psychotherapy as well. Chatbots/conversational agents have been developed with AI technology. These programs have the ability to maintain a dialogue similar to human-human interaction. In other words, a non-human entity exhibits language processing abilities similar to and sometimes even exceeding those of humans (Haber et al. 2024). In current versions of AI systems that use large language models (LLM) (e.g. ChatGPT 4.0), it has become possible to conduct a dialog with the AI (OpenAI 2023). It is stated that chatbots, like human therapists, can communicate with users to help them recognize their emotions and thought patterns and offer coping methods (Meadi et al. 2024). Companies developing AI-supported therapeutic applications aim to make these applications similar to face-to-face therapies with tools such as chatbots and virtual reality therapies (Holohan and Fiske 2021).

With these developments, AI applications are becoming more and more common in psychotherapy. It is claimed that AI can ensure the active participation of patients in the recovery process thanks to the personalized psychotherapy that it will conduct by considering the clinical, cultural, and personal characteristics of the patient with the information obtained from large data sets (Haber et al. 2024). The ability to apply evidence-based therapies such as cognitive behavioral therapy (CBT), and to create a pluralistic environment that integrates approaches such as CBT and psychoanalytic therapy with new and developing therapy modalities are presented as other advantages that AI provides to psychotherapy (Haber et al. 2024). In addition, augmented reality and virtual reality technologies offer opportunities for exposure therapy by creating controlled and safe scenarios (Ayhan 2023). The groups that are thought to benefit the most from AI-assisted psychotherapy are people who have problems in accessing treatment due to fear of stigmatization, economic difficulties, lack of therapists, and vulnerable groups (Fiske et al. 2020).

AI also has challenges, such as deficits that may be encountered in clinical practice and a lack of standardized practice. This raises ethical concerns, too. The first question that comes to mind data security. Especially in the psychotherapy process, when a lot of sensitive and private information about individuals is stored in the AI database, patients' personal data may become available for unauthorized data use or detailed analysis of the data without explicit consent. How to protect this data is a serious concern. On the other hand, it is another concern if the data is collected under the control of companies involved in AI development. This could lead to prioritizing commercial interests and jeopardize service quality (Zajko

2023). These issues point to the risk of violating ethical principles such as security and privacy.

There is a risk that AI may produce biased or erroneous results (Panch et al. 2019). It is also known that AI can generate fabricated responses that appear to be persuasive and consistent within the text, but are completely independent of user input or previous context. This is referred to as AI hallucination or confabulation (Özer 2024). In other words, the responses produced by the AI may be meaningless or inaccurate, even if they seem plausible. As a result, the patient can be formulated according to the wrong results and psychotherapy can be conducted with erroneous or fabricated responses.

Although LLMs appear neutral and objective, they learn a set of value systems shaped by culture, gender, race, ethnicity, socio-economic factors, and so on. The transparency of this learning and the exact formation of the value system remain unclear (Hadar-Shoval et al. 2024). Consequently, AI-supported therapies may lack homogeneity, and non-transparent processes could result in varied applications. Additionally, the mechanism by which these differences and the therapy process are monitored is not well-defined. Concerns may escalate further if therapists are employed by AI companies, as conflicts of interest could complicate the evaluation of AI-based psychotherapy, the open sharing of test outcomes, and addressing potential drawbacks in favor of the patient. The absence of legal regulations and ethical guidelines governing AI development and use in psychotherapy further exacerbates these concerns, potentially violating the ethical principles of nonmaleficence and beneficence.

Another issue is that evidence regarding the effectiveness and reliability of using AI in psychotherapy is still insufficient. Widely used chatbots are often introduced without being based on empirical evidence. Due to the lack of sufficient research in this area, caution is necessary when making general conclusions.

It does not yet seem plausible for AI to function as a stand-alone therapist, especially when it comes to psychoanalytic psychotherapies. Sedlakova and Trachsel (2022) argue that AI's role in psychotherapy is limited to conveying information based on the data it gathers and that it would fall short in more advanced functions. According to the authors, if users' expectations of AI include establishing an authentic dialogue, gaining new self-understanding, insights, or feeling human closeness, AI may not be sufficient for these purposes, and users may be misled. This misconception could lead to patients being unable to independently develop new attitudes and behavior patterns about themselves, thereby reducing their autonomy (Sedlakova and Trachsel 2022).

The introduction of AI into psychotherapy poses several issues related to the basic elements of psychotherapy. Chatbots do not have emotions; hence, they lack the ability to display genuine empathy. Although some chatbot responses are interpreted as empathic by users (Inkster et al. 2018), AI responses are learned and chatbots cannot show genuine empathy. Due to the lack of empathy, one of the most essential qualities in a therapist, chatbots are thought to have a higher potential to harm patients and a lower ability to benefit patients compared to human therapists (Meadi et al. 2024). Human-human interaction in psychotherapy, i.e. the human touch, strengthens the therapeutic relationship. The therapeutic relationship, whose healing power in psychotherapy is undeniable, will not be established sufficiently with chatbots due to the problem of empathy.

The significance of establishing the therapeutic relationship with software instead of a human is an important topic that needs to be discussed. The concept of transference, which is central to therapy, may also change with the introduction of AI into psychotherapy. Transference has a place and value in all forms of psychotherapy collaboration, including cognitive-behavioral therapy (CBT) (Prasko et al. 2022). Transference should be a consideration in every psychotherapy involving AI.

In a study (2021), Holohan and Fiske raised many important questions about this issue: "Does transference occur with the introduction of AI into psychotherapy? If so, how does it happen and how is the subsequent therapeutic relationship and practice affected? How might the patient relate to the chatbot? Through which words, behaviors, attitudes can the development of transference be monitored? How can transference be understood, explained and addressed in AI-assisted therapy? How can developers and engineers enable the use of transference in the design of AI-assisted therapeutic applications? Is this transference as we are familiar with it? Or is there some other kind of relationship that looks like transference but is somehow different?" There is no doubt that it is necessary to reflect on these questions and gather evidence-based data to answer them. However, there are no publications that study transference in AI-assisted psychotherapy. Nevertheless, there is evidence in some studies that there is a relationship between a human and a chatbot similar to a human-human relationship (Fulmer et al. 2018). This indicates that transference can develop. The development of transference in psychotherapy is inevitable. The use and handling of transference is essential, especially for psychoanalytic psychotherapy. However, it is not easy to predict how the AI can realize the developing transference. Since it has no emotions and cannot recognize emotions, the chatbot can only analyze the transference according to the patient's utterances and conclude about the transference. However, this is not a genuine understanding and interpretation of the transference. On the other hand, it is unclear what the patient's understanding of the developing transference towards the relationship with AI signifies in

their daily life and whether it will benefit the patient. As a matter of fact, it is not known how close these feelings are to transference in the classical sense.

Although the companies market the 24/7 availability of chatbots as as positive feature, it can lead to serious issues in psychotherapy. Patients who can access the chatbot at any moment may find it difficult to adhere to the therapy framework and respect boundaries. Those seeking help from AI for every issue may struggle to develop their coping skills. The constant presence of a "therapist" may fuel the patient's savior fantasies. Due to the impact of the transference they develop, patients may establish a dependent relationship with AI and may trust the chatbot unconditionally, without any questioning (Sedlakova and Trachsel 2022).

In psychotherapy, where there is transference, there is also countertransference. Considering that AI lacks the ability to generate and recognize emotions but can only imitate them (Weber-Guskar 2021), problems with understanding countertransference can be expected in psychotherapies using AI. The inability to use countertransference by AI makes it difficult to understand the patient and to identify the relationship he/she establishes with others in his/her daily life and the way he/she makes others feel. This means a setback in the therapy process.

Being a therapist is not the only role AI can take in psychotherapy. AI applications can also appear as an auxiliary element in psychotherapy. Sedlakova and Trachsel (2022). Tried to define AI in psychotherapy in two poles. The first one is to define AI as a therapeutic tool that does not deal with the state of mind and intention of the patient, does not connect, and plays an auxiliary role in therapist-patient communication. The second one defines AI as an active agent that thinks about situations, builds relationships, and learns (or imitates) empathy and emotions. According to the authors, the place of AI in psychotherapy is neither of these. It is not possible to reduce AI to a tool because AI can communicate and relate to patients. Moreover, using AI only as a tool may mean ignoring its possible wide use and potential benefits in psychotherapy. On the other hand, it is not possible to define AI as an effective actor because it lacks human abilities such as empathy, understanding and conceptualizing the other's state of mind, purposeful application, and bonding. Additionally, using AI as an effective actor means placing the responsibilities assigned to human therapists onto AI, with its current level of advancement. However, AI does not have the competence to assume this responsibility yet, and doing so could lead to many ethical issues. Instead of this distinction, the authors propose a hybrid use of AI as both a tool and an actor in therapy (Sedlakova and Trachsel 2022). Although these views of the authors are plausible, new studies as well as ethical and legal regulations are needed to find the balance of using AI as a tool and as an effective actor.

Despite all these problems and possible negativities, it is clear that AI has a potential to make psychotherapy more accessible. Therefore, another question comes to mind: "Should AI-assisted psychotherapies be made available to patients who cannot access psychotherapy or should such services be avoided until the data are sufficient?" This early question will inevitably come up as AI develops and its application areas increase.

In conclusion, it is clear that AI has the potential to bring about significant changes and advancements in the practice of psychotherapy in the future, as well as to make psychotherapy more accessible. In addition to the many positive effects that may result from the incorporation of AI into psychotherapy, there may also be uncertainties and problems, as outlined in the paper, which may result in patient harm. All these issues are still at the level of speculation. As reflections on the subject deepen, many other issues may come to the minds of mental health professionals. It can be predicted that many dilemmas will arise with the inclusion of AI not only in psychotherapy practice but also in training and supervision processes. In this article, we have mainly tried to address the practical and ethical aspects of using AI in individual psychotherapy. It can be easily assumed that the addition of AI to various practices such as couple therapy, sexual therapy, group therapy, and counseling interviews will pave the way for the development of other negativities and ethical problems specific to these practices.

There is a need for empirical studies and the reporting of patients' positive and negative experiences regarding the validity of all the speculations and how they can be addressed. On the other hand, to achieve the most positive outcomes, software developers, mental health professionals, and experts in ethics and law need to work collaboratively to improve existing AI tools. Legal regulations should be developed to ensure the well-being of patients by establishing laws and guidelines that respect patient rights and ethical principles. Given that the practice of psychotherapy and interaction with patients may vary according to cultural differences, it will not be enough to follow international literature and regulations and shape our practices accordingly. Teams should also be formed in Turkey, and studies should be conducted regarding the use of AI in psychotherapy, which is inevitable to enter clinical practice in the future.

REFERENCES

Ayhan Y (2023) The Impact of Artificial Intelligence on Psychiatry: Benefits and Concerns-An assay from a disputed 'author.' Turk Psikiyatri Derg 34: 65-7

Fiske A, Henningsen P, Buyx A (2020) The implications of embodied artificial intelligence in mental health- care for digital wellbeing. Ethics of digital well-being, C Burr and L Floridi (Ed), 1st edition, Cham, Springer International Publishing, s: 207–19.

Gürcan A, Pereira-Sanchez V, Pinto da Costa M et al. (2024) Artificial intelligence innovations in psychiatry: Global perspective from early career psychiatrists. Turk Psikiyatri Derg 35: 83-4.

Haber Y, Levkovich I, Hadar-Shoval D et al. (2024) The Artificial Third: A Broad View of the Effects of Introducing Generative Artificial Intelligence on Psychotherapy. JMIR Ment 11: e54781.

Hadar-Shoval D, Asraf K, Mizrachi Y et al. (2024) Assessing the alignment of large language models with human values for mental health integration: cross-sectional study using Schwartz's theory of basic values. JMIR Ment 11: e55988.

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. Front Psychol 12: 720476.

Inkster B, Sarda S, Subramanian V (2018) An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Ment 6: e12106

Meadi MR, Bernstein JS, Batelaan N et al. (2024) Does a lack of emotions make chatbots unfit to be psychotherapists? Bioethics 38: 503–10.

OpenAI (2023) ChatGPT-4. Accessed October 5, 2024. [Large language model]. Computer software.

Özer M (2024) Is artifical intelligence hallucinating? Turk Psikiyatri Derg 35:333-5. https://doi.org/10.5080/u27587

Panch T, Mattie H, Atun R (2019) Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 9: 010318.

Prasko J, Ociskova M, Vanek J et al. (2022) Managing Transference and Countertransference in Cognitive Behavioral Supervision: Theoretical Framework and Clinical Application. Psychol Res Behav Manag 11: 2129-55.

Sedlakova J, Trachsel M (2022) Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent? Am J Bioeth 23: 4-13

Weber-Guskar E (2021) How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners. Ethics Inf Technol 23: 601–10

Zajko M (2023) Automated Government Benefits and Welfare Surveillance. Surveill Soc 21: 246-58.

Hayri Can ÖZDEN

Received: 06.10.2024, **Accepted:** 10.10.2024, **Available Online Date:** 14.10.2024

Psychiatrist, Hatay Samandağ State Hospital, Department of Psychiatry, Hatay, Turkey.

Dr. Hayri Can Özden, e-mail: hayricanozden@hacettepe.edu.tr

https://doi.org/10.5080/u27603