ADVANTAGES AND DISADVANTAGES OF AI CONVERSATIONAL AGENTS IN COGNITIVE THERAPY

Valentina Neacșu,

"Titu Maiorescu" University, Faculty of Psychology, Bucharest

DOI: https://doi.org/10.66793/titup19proceeding10

***ABSTRACT*:**

AI-based conversational agents (CAs) offer significant benefits in therapy, including 24/7 availability, low cost, and increased accessibility, making mental health support more reachable for underserved populations. These tools can provide timely assistance and facilitate psychoeducation, especially in crisis situations. However, their use raises important ethical and privacy concerns, such as data security, algorithmic bias, and the lack of emotional depth in AI interactions. Additionally, the potential for AI-induced psychosis in vulnerable individuals underscores the need for careful regulation. This article examines both the advantages and drawbacks of AI-based CAs in therapy, highlighting the need for ethical guidelines and safeguards to ensure that AI enhances, rather than replaces, human therapeutic care.

Keywords: *Artifical intelligence, AI conversational agents, chatbots, data privacy, mental health*

1. INTRODUCTION

With the rapid rise and constant evolution of artificial intelligence (AI) models, covering multiple areas of study, the intersection of AI and mental health raises important questions and issues. AI driven tools, like chatbots or even virtual therapists, promise users real-time support and validation, often times for free or at a very low cost.

However, can such tools handle sensitive data safely? Can they imitate the nuanced understanding and years of experience of a human therapist? Is it possible for AI to offer the same level of effectiveness compared to traditional, face-to-face therapy?

By examining both the positive and negative aspects of incorporating these emerging technologies into cognitive therapy practices, this article aims to offer a balanced perspective on using AI as a tool, as opposed to replacing the traditional therapeutic process entirely. As we move forward, it is imperative to understand the pitfalls and potential of AI tools to make sure it is used responsibly, ethically and effectively in therapy.

2. RISE OF AI IN MENTAL HEALTH

A particularly promising form of digital mental health intervention takes the form of mobile apps, which can be easily installed on a person's device. Recent estimates show that over 10,000 mental health apps are now available to consumers[1]. Mobile apps offer the convenience of accessing mental health support anytime and anywhere. Most mental health-oriented apps are either moderated by mental health professionals or are entirely autonomous conversational agents [1],[3].

Mehta et al.[1] defines artificial intelligence (AI) therapy as a digital and fully automated, mobile, psychological treatment program that uses a conversational interface to deliver just-in-time adaptive interventions. The 3 key features that set AI therapy apart from traditional digital intervention approaches are (1) the use of a conversational (chatbot) interface, (2) inclusion of just-in-time interventions, and (3) adaptation and personalization.

Conversational agents (CAs), or chatbots, have demonstrated considerable potential in mental health care. They can aid in diagnosis and consultations, offer psychoeducation, and provide treatment options, while also contributing to social support and enhancing mental resilience [2],[3].

However, most of these CAs still function on rule-based systems, which depend on predefined scripts or decision trees to interact with users. The rule based models of CAs are limited due to their inability to fully understand user context and intent. Recent advancements in AI, particularly natural language processing (NLP) and generative AI, have paved the way for a new generation of AI driven CAs. They can now process more complex information, enabling them to provide more personalized, adaptive and nuanced responses compared to the rule based CAs [3].

Li et al. performed a systematic review and meta-analysis, showing that AI-powered conversational agents can be effective in enhancing mental health and well-being. Their research emphasizes that chatbots offer timely assistance and tailored recommendations, leading to notable reductions in anxiety and improvements in psychoemotional well-being, particularly among younger users [2].

In her recent study, Spytska compared the effectiveness between CAs and traditional therapy for two groups that had anxiety disorders from being in active war zones. One group had access to the Friend chatbot for daily support, while the other had 3 1-hour long therapy sessions per week [4].

Both groups experienced significant reductions in anxiety levels. The control group, which received traditional therapy, showed a 45% reduction on the Hamilton scale and a 50% reduction on the Beck scale. In comparison, the chatbot group saw reductions of 30% and 35%, respectively. While the chatbot offered accessible, immediate support, traditional therapy was more effective due to the emotional depth and flexibility provided by human therapists. The chatbot proved especially valuable in crisis situations, where access to therapists was limited, highlighting its potential for scalability and availability. However, its emotional engagement was noticeably lower than that of in-person therapy [4].

3. THE ADVANTAGES OF USING AI IN COGNITIVE THERAPY

The main advantages of AI based technologies, generative or CAs, are especially linked to ease of access.

While traditional face-to-face therapy faces both structural barriers (availability and financial situations) and behavioral ones (patients feel the need to handle problems independently and privately), AI-based tools can be accessed from anywhere and at any time, offering support outside of regular therapy hours. This is particularly beneficial for individuals in remote areas or those with limited access to mental health professionals [1],[3],[4].

AI based CAs and generative models can provide cognitive therapy at scale, reaching many individuals simultaneously, their lack of bias and judgement making clients less fearful, compared to human interactions.

Barring that, AI programs can offer support and guidance in multiple languages and it can lower the overall cost of therapy by providing supplementary support, making it an attractive alternative for impoverished communities [2],[4],[5].

Users may feel more comfortable sharing sensitive personal information with AI, knowing they are not speaking to a human who might share the information. This is especially true with younger users who are more at ease using technology [1],[6].

However, this trust is a double-edged sword, as shown below.

4. THE DISADVANTAGES OF USING AI IN COGNITIVE THERAPY

Despite their benefits, AI-based conversational agents (CAs) come with risks, including privacy concerns, biases, and safety issues. Their unpredictable nature may result in inaccurate or harmful outcomes, potentially leading to unintended negative consequences [5].

To ensure the safe and effective integration of AI-based CAs into mental health care, it is crucial to conduct a thorough review of the current research on their use in mental health support and treatment. This will help inform healthcare providers, technology developers, policymakers, and the public about the evidence supporting the effectiveness of these technologies, while also identifying challenges and areas that require further investigation [12].

A. Inability to fully understand emotions and lack of accountability

While AI can simulate conversations, it cannot truly understand human emotions or provide the empathy that a trained therapist can offer, which is essential for effective cognitive therapy [15].

Some users may feel more isolated or detached when interacting with AI, as opposed to engaging with a human therapist who can offer emotional support and understanding.

AI lacks the ethical accountability that human therapists have. If AI makes a mistake or gives harmful advice, it may not be clear who is responsible for the consequences [3],[6].

As AI emerges as one of the most innovative and rapidly evolving tools in psychological and psychiatric research and treatment, existing legal and ethical frameworks often fail to keep pace with these developments. Instead of offering proactive regulatory guidance, there is a risk that the gaps between the application of AI technologies and ethical standards will only be addressed after harm has already occurred [6],[9].

Sadly, recent events have already showed the dark side of using relying on CAs as sole therapeutic practices, showing self-harm or suicide attempts following use increased ,, as well as ”AI psychosis“, an unofficial diagnostic describing symptoms triggered by prolonged use [7],[8].

In his study, Preda showcases the leading causes of negative mental effects of AI chatbot use. CAs are more likely to mirror user responses rather than challenge them, a result of training designed to prioritize agreement

over accuracy. Additionally, preference-optimized models reinforce users' views, making them unable to distinguish or challenge delusional beliefs [9].

Persistent memory features, intended to enhance user experience, may unintentionally reinforce these delusions by carrying paranoid or grandiose themes across sessions., Lastly, individuals with certain characteristics, such as those at risk for psychosis, those with autistic traits, or those experiencing social isolation, may be more vulnerable to adverse AI-induced effects, including AI-triggered psychosis [9],[10].

B. Ethical and privacy concerns

From an ethical standpoint, the use of AI in mental health offers several potential benefits, including new treatment methods, improved access to hard-to-reach populations, enhanced patient engagement, and the ability to free up time for healthcare providers [12],[13].

However, there are also significant ethical concerns, such as harm prevention, data privacy issues, and the lack of clear guidelines for developing AI applications, integrating them into clinical settings, and training healthcare professionals [9],[10],[14],[15].

Fiske et al. highlights the fact that there are gaps in ethical and regulatory frameworks, and the rising risk of misuse—such as replacing traditional services with AI, which could exacerbate existing health disparities. Specific challenges in implementing AI tools in the therapeutic process include: assessing risk, managing referrals and supervision, ensuring patient autonomy is respected and protected, maintaining transparency in algorithmic decision-making, and addressing concerns about the long-term impact of these technologies for patients [5].

C. Data security

Storing and processing vast amounts of sensitive mental health and personal data raises questions about how securely that information is handled by corporations.

In order to function, AI chatbots require the collection and storage of large amounts of personal data, raising significant concerns about data security and privacy. For AI models to function effectively, they rely on continuous machine learning, which involves feeding vast quantities of data into the chatbot's databases. If this data includes sensitive patient or business information, it becomes part of the dataset used by the chatbot in future interactions. As a result, this data could be exposed to both intended and unintended audiences and potentially used for unauthorized purposes [3],[4],[5],[10].

A recent study reveals that leading AI companies are pulling user conversations for training, highlighting privacy risks and a growing need for clearer policies. King et al. examined AI developers' privacy policies and identified several causes for concern, including long data retention periods, training on underage users' data, and a general lack of transparency and accountability in developers' privacy practices. The authors recommend that AI users should reconsider the information they share in CAs and, whenever possible, opt out of having their data used for training [11],[14].

D. Dependence on technology

AI tools are not immune to technical glitches, outages, or algorithmic biases, which could disrupt therapy sessions and undermine user trust in the process [5],[10].

Moreover, access to AI-driven therapy might be limited for individuals who do not have reliable internet access or compatible devices to engage with the technology [4].

5. CONCLUSIONS

AI-based conversational agents (CAs) and generative models hold great promise in revolutionizing therapy, offering significant advantages such as ease of access, affordability, and round-the-clock availability. The ability to access therapeutic support at any time can be especially beneficial for individuals in remote areas, those with limited access to traditional therapy, or those who need immediate assistance during moments of crisis. Furthermore, the low cost of these interventions could help bridge the gap for people who may otherwise be unable to afford ongoing mental health care. As such, AI could play a pivotal role in improving mental health outcomes for a wide range of users.

However, the use of AI in therapy also raises several ethical and privacy concerns that cannot be overlooked. One of the most pressing issues is the collection and handling of sensitive personal data, which could expose users to potential security breaches, data misuse, or unauthorized surveillance. The use of large datasets to train AI models also raises concerns about bias, as these models can inadvertently reinforce harmful stereotypes or fail to recognize the complexities of human psychology. In addition, while AI can mirror human-like interactions, it lacks the nuanced understanding and emotional empathy that human therapists provide, which are essential for addressing deeper psychological issues.

As AI technologies continue to advance, it is crucial that regulatory frameworks, ethical guidelines, and safeguards evolve alongside these tools to prevent harm and ensure that AI remains a supportive complement to, rather than a replacement for, human care. Balancing the benefits of accessibility, affordability, and convenience

with the need for ethical accountability and privacy protection will be key to ensuring that AI-based CAs enhance, rather than undermine, the therapeutic experience

REFERENCES:

[1] *A. Mehta et al., "Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (Youper): Longitudinal observational study," Journal of Medical Internet Research, vol. 23, no. 6, Jun. 2021. doi:10.2196/26771*

[2] *H. Li, R. Zhang, Y.-C. Lee, R. E. Kraut, and D. C. Mohr, "Systematic Review and meta-analysis of AI-based conversational agents for promoting mental health and well-being," npj Digital Medicine, vol. 6, no. 1, Dec. 2023. doi:10.1038/s41746-023-00979-5*

[3] *A. I. Jabir et al., "Evaluating conversational agents for Mental Health: Scoping Review of outcomes and outcome measurement instruments," Journal of Medical Internet Research, vol. 25, Apr. 2023. doi:10.2196/44548*

[4] *L. Spytska, "The use of Artificial Intelligence in psychotherapy: Development of intelligent therapeutic systems," BMC Psychology, vol. 13, no. 1, Feb. 2025. doi:10.1186/s40359-025-02491-9*

[5] *A. Fiske, P. Henningsen, and A. Buyx, "Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy," Journal of Medical Internet Research, vol. 21, no. 5, May 2019. doi:10.2196/13216*

[6] *T. Vial and A. Almon, "Artificial Intelligence in mental health therapy for children and adolescents," JAMA Pediatrics, vol. 177, no. 12, p. 1251, Dec. 2023. doi:10.1001/jamapediatrics.2023.4212*

[7] *K. Payne, "An AI chatbot pushed a teen to kill himself, a lawsuit against its creator alleges," Associated Press News, Oct. 26, 2024*

[8] *S. John, "Why AI companions and young people can make for a dangerous mix," Stanford Medicine News Center, Aug. 27, 2025*

[9] *A. Preda, "Special report: Ai-induced psychosis: A new frontier in Mental Health," Psychiatric News, vol. 60, no. 10, Oct. 2025. doi:10.1176/appi.pn.2025.10.10.5*

[10] *J. Sedlakova and M. Trachsel, "Conversational Artificial Intelligence in psychotherapy: A new therapeutic tool or agent?," The American Journal of Bioethics, vol. 23, no. 5, pp. 4–13, Apr. 2022. doi:10.1080/15265161.2022.2048739*

[11] *J. King, K. Klyman, E. Capstick, T. Saade, and V. Hsieh, "User privacy and large language models: An Analysis of Frontier Developers' privacy policies," Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, vol. 8, no. 2, pp. 1465–1477, Oct. 2025. doi:10.1609/aies.v8i2.36646*

[12] *J. Prescott and S. Barnes, "Artificial Intelligence Positive Psychology and therapy," Counselling and Psychotherapy Research, vol. 24, no. 3, pp. 843–845, Jun. 2024. doi:10.1002/capr.12784*

[13] *P. B. Ooi and G. Wilkinson, "Enhancing ethical codes with artificial intelligence governance – a growing necessity for the adoption of Generative AI in counselling," British Journal of Guidance & Counselling, vol. 53, no. 1, pp. 66–80, Jul. 2024. doi:10.1080/03069885.2024.2373180*

[14] *A. Alhuwaydi, "Exploring the role of Artificial Intelligence in mental healthcare: Current trends and Future Directions – A narrative review for a comprehensive insight," Risk Management and Healthcare Policy, vol. Volume 17, pp. 1339–1348, May 2024. doi:10.2147/rmhp.s461562*

[15] *D. Richards, "Artificial Intelligence and psychotherapy: A counterpoint," Counselling and Psychotherapy Research, vol. 25, no. 1, Apr. 2024. doi:10.1002/capr.12758*