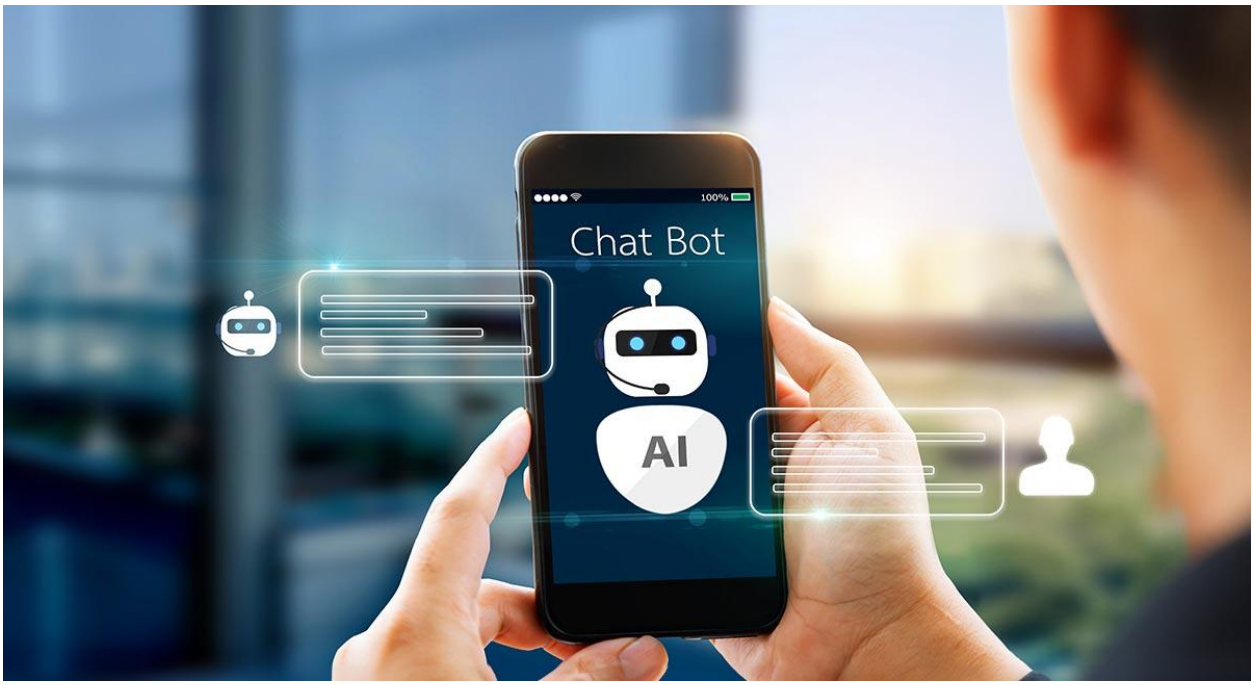


The Responsible Integration of AI in Mental Health Care

Submitted to

Indy Recker
University of Oregon
Eugene, OR

May 27, 2025

by

Alexia Crawford

This report examines the ethical implications of using artificial intelligence in mental health care. It reviews both the benefits and the challenges, including data privacy, loss of human empathy, and treatment complications. This section covers concerns of AI integration, emphasizing the importance of human oversight, informed consent, and patient autonomy.

Table of Contents

List of Figures

1.0 Introduction

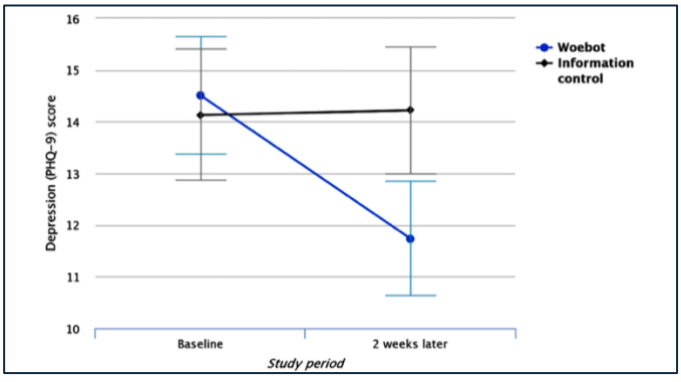
As artificial intelligence (AI) continues to transform the healthcare industry, its integration into mental health services, especially amongst adolescents, has introduced many new possibilities, but also at the same time raising ethical questions. While AI has been widely used in diagnostics, imaging, and more, its application in mental health care is newer. Because it is newer, it needs to be looked at with a different lens, which looks at all aspects, including ethical concerns. Adolescents represent a vulnerable population as they deal with developmental, emotional, and social challenges while also growing up in a digital age. This report is part of a bigger report that focuses on AI in health care, as opposed to mental health care. Issues such as privacy, effectiveness, and AI's emotional intelligence will be analyzed in regard to AI in mental health care. By focusing on this area of mental health care, the section contributes to the report's broader goal of identifying pathways for implementing AI technologies across the healthcare system, and the issues it is currently facing. This section is intended for representatives, healthcare providers, and technology developers who are working on merging AI and mental health care. By highlighting multiple challenges, it aims to support the creation of AI tailored to the needs of users.

2.0 Accessibility of AI Therapy

There are multiple AI-powered therapeutic tools out right now such as Woebot that are increasingly being recognized for their ability to expand access to mental health care. This access is important especially for people and groups who may face barriers to traditional therapy. One of the most pressing issues in mental health today is the huge demand for services, with nearly half of psychologists reporting that they are unable to meet the demand for treatment and over 70% maintaining longer waitlists than before the pandemic (Bethune, 2022). In this context, AI therapy is as a critical supplementary resource, needed to help all these people. In a randomized controlled trial, Woebot was shown to "increase engagement with most individuals using the bot nearly every day," and participants "viewed it more favorably than the information-only comparison" (Fitzpatrick et al., 2017). As shown in figure 1, the same study demonstrated a statistically significant reduction in symptoms of depression and anxiety.

Figure 1

Change in Depression by Group



By Fitzpatrick et al., 2017

Change in depression scores over two weeks, comparing Woebot users to a control group.

Along with addressing availability, these AI tools can also address the comfort level of the people using them. In particular, adolescents prefer anonymous and tech-based support, as they often feel misunderstood by adults (Alfono et al., 2024). This can help reduce the stigma often associated with in person therapy. This supports innovation efforts by other companies such as Johnson & Johnson. Johnson & Johnson's AI tools are being applied in many ways. They are applying these tools in diagnostics and patient monitoring, and it proves the growing potential for AI to personalize care across many healthcare fields (Johnson & Johnson, 2024). As technology and AI continue to improve and become more implemented, they need to meet mental health needs that aren't currently being met. This is especially important amongst adolescents because they grew up in a digital world. These tools can also "expand access to patients who might otherwise not be able to access treatment, particularly patients from underserved communities, such as those living in rural areas and communities of color" (Bethune, 2022). By integrating AI solutions into mental health care, the strain on human therapists and traditional therapy will lessen and it will ensure more equitable support for those who need it.

3.0 Ethical Challenges: Privacy, Consent, and Autonomy

As AI tools become more integrated into mental health care, especially amongst adolescents, many ethical concerns arise. These concerns regard privacy, informed consent, and personal autonomy. While these AI tools can offer more support and convenience, they also bring up risks that can't be overlooked or brushed aside.

One of the primary ethical issues is data privacy. AI mental health tools such as Woebot use user inputs to provide responses which involves storing large amounts of sensitive mental health information. In traditional therapy, notes are protected under strict confidentiality policies. AI therapy must rely on online storage, and the information has to be stored in the cloud, which can lead to some issues and vulnerabilities. A critical question is raised, "Where will the data from these conversations be stored? Who will be responsible for the clinical records, the conversation history stored in the cloud?" (Alfano et al., 2024). They go on to explain that the companies managing this data could shut down or even sell the data to third parties such as employers. To developers who are implementing these systems, this raises an urgent call for clarity around who manages and protects data. This would be a major breach of privacy and trust, further raising privacy concerns.

While these issues are already prevalent, they are even more prevalent in certain groups. The first being adolescents. They often lack a full understanding of the long-term effects of consenting to AI driven mental health care. While they may be more comfortable with technology than other age groups, they still might not understand the privacy policies or other important fine texts. Communities of color are an example of another group that may be affected by problems such as these. They already face disparities in mental health care and access. People

of color are more likely to be underserved and misdiagnosed in traditional therapy. AI mental health care tools can be helpful when serving these communities, especially if they are in rural or under-resourced areas (Bethune 2022). Again, without strong regulations, these tools risk worsening existing inequalities within communities of color and mental health care.

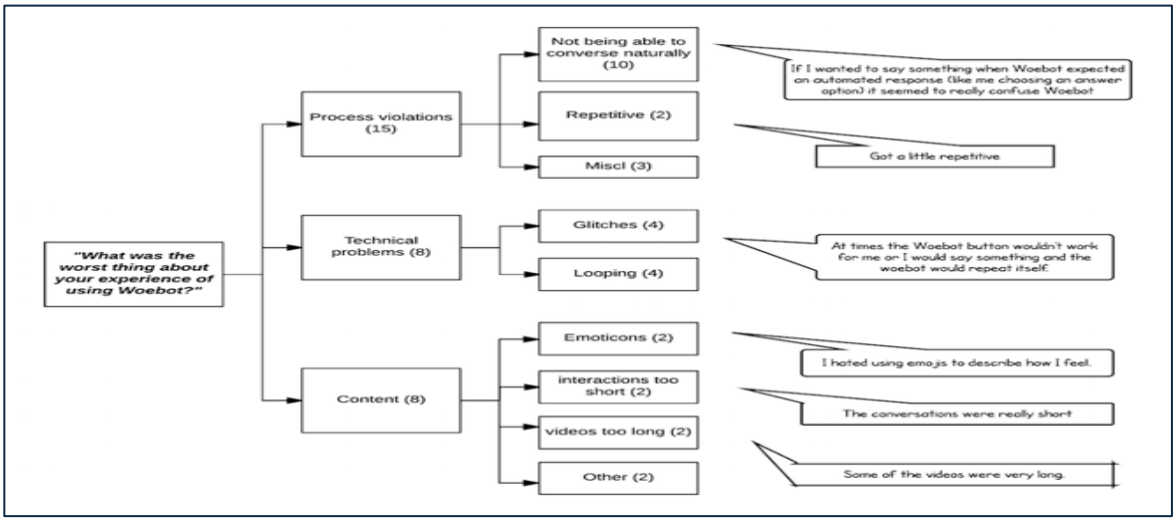
While AI offers opportunities to expand access, it must be developed with careful attention to ethical and privacy concerns. These need to protect vulnerable users, including adolescents and people of color. Making sure data practices are transparent is essential to building trust and equitable mental health care tools.

4.0 Humans vs. AI: Empathy and Depth

While AI therapy tools like Woebot have shown success in reducing symptoms of depression and anxiety, they have certain limitations that show up when they are compared to traditional, human, therapists. These limitations are apparent in emotional depth and connection. One main concern raised by researchers is that even though AI bots can replicate the structure of therapy dialogue, they lack the ability to form emotional connections an symbolic interpretation. One study says, “The first limitation relates to the machine’s inability to grasp the emotional, symbolic, relational, and anthropological dimensions of the data it is meant to interact with,” (Alfano et al., 2024). A human therapist specializes in recognizing unspoken behaviors, tone, and can respond empathetically, while current AI mental health care tools lack those skills.

Even though these chat bots are designed to imitate human connection, these conversations and interactions are superficial. For example, that “the bot’s conversational style was modeled on human clinical decision making and the dynamics of social discourse,” yet they acknowledge the risk of users feeling the responses as mechanical or detached (Fitzpatrick et al., 2017). The absence of human empathy creates a risk of surface level or misinterpreted therapeutic interactions which could delay long-term usage. This issue could potentially cause harm to the user if they are feeling misunderstood or like they aren’t being listened to.

Figure #2
Group's Least Favorite Things About Woebot



Note. By Fitzpatrick et al. (2017). Reported issues with Woebot use, including technical glitches, unnatural responses, and content concerns.

Despite these issues, these AI chat bots have still shown positive outcomes in therapeutic settings. Fitzpatrick et al. report that "those in the Woebot group significantly reduced their symptoms of depression over the study period" in a controlled trial. These results show that Woebot, the AI powered therapy chat bot, helped certain users feel less anxious and depressed. However, in another study, it was warned to not over rely on AI in mental health cases (Lee et al., 2021). They argue that while AI excels in certain cognitive functions, it still can't replicate human emotional intelligence. Emotional intelligence is crucial to receiving the best care possible. While AI may serve as a complement to traditional therapy, it lacks certain key aspects of human interaction

5.0 Risks and Long-Term Effects

Although these tools have shown positive progress with patients, there are issues that persist amongst their long term effects, especially among adolescents. One of the biggest issues is user retention. One study says "One of the most significant challenges is related to dropouts, which are associated with the ease of access and familiarity with the digital medium," (Alfano et al., 2024). Users who want immediate feedback and results may be disappointed when they can't immediately see their progress or have someone telling them how their progress is going. This often leads to quitting the treatment and reporting that it is ineffective. This has long term implications. This same study explains that even when many users continue to feel distressed,

they don't return to treatment for up to five years (Alfano et al., 2024). The risk of disengagement is especially concerning for adolescents. They are still developing their emotional regulation skills and often struggle with the stigma that surrounds therapy. If the AI chat bot is their first attempt at getting help, and it doesn't meet their expectations, they could be discouraged from getting help altogether.

More concerns arrive when considering other mental health challenges. According to the American Psychological Association (Bethune, 2022), nearly half of psychologists have no room for new patients, and over 70% report longer waitlists than pre-pandemic. If users aren't satisfied with their AI mental health care, they might not have other options. Traditional therapists don't have space for new patients. This is another reason why AI tools need to be designed for long term usage, not just short-term engagement. Without tools to recognize the disengagement risks, these AI mental health care tools might form a cycle of unmet needs. Developers and mental health providers must recognize these behavioral patterns and intervene early. As Alfano et al. note, early dropout is not only a temporary setback, but a strong predictor of long-term disengagement. While AI tools can expand access, their design must be sensitive to long-term behavioral patterns, or else short-term dropout will lead to negative consequences.

6.0 Future Applications

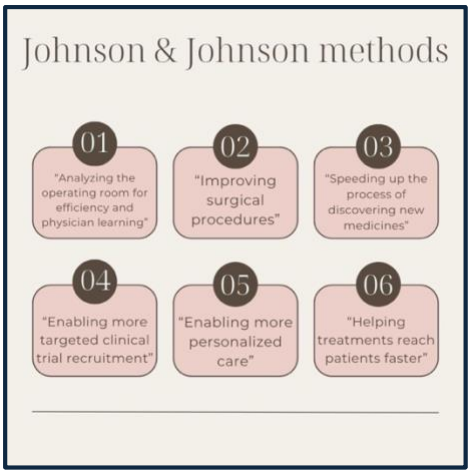
AI is evolving from a supportive digital tool to an essential tool in mental health care. Because it is modeled after human intelligence, "AI can accomplish various concrete tasks far more quickly than a human by replicating discrete human intelligence skills such as processing speed, memory, quantitative reasoning, visuospatial ability, auditory processing, and comprehension knowledge," (Lee et al., 2021). This allows AI tools to analyze large amounts of data to detect concerns such as depressive language or suicidal thoughts. This data can include mood logs, messages, vocabulary, and more. As these AI tools act as a warning system, they can help human therapists in their work. AI tools like Woebot have already demonstrated this capability on a smaller scale. As these tools continue to develop, it's possible that AI tools will be the standard daily care with human intervention as necessary. This hybrid model can improve care quality and equity across groups.

However, there are issues that still remain. Lee et al. warn that "the underlying biological processes of psychiatric disorders are still poorly understood," which limits how effectively AI can interpret symptoms or personalize treatment without risking misjudgment. To limit how often this happens, these tools need to collect data and respond with emotional intelligence and ethical standards. To increase understanding and follow up engagement, these AI tools could include features like follow up messages. Another option could be the choice to switch to live therapy when it feels needed. It is also important to set standards at the start. Users, particularly

adolescents, might expect immediate results and then disengage when those results aren't met. These tools have to set a precedent that it's a gradual process and normalize the ups and downs of mental health care.

AI can be used in many health care settings and in many ways. Johnson & Johnson's work and how AI can be implemented as an assistant, from optimizing surgical procedures to targeting treatments. Their work serves as a model for mental health AI care by combining efficiency with personalization and professional oversight (Bethune 2022). AI's future in mental health lies in its ability to assist, not replace, human care while making therapy more responsive and accessible to user needs.

Figure 3
Johnson & Johnson AI Implementation methods



Note. By Alexia Crawford, 2025, How Johnson & Johnson implements AI tools.

7.0 Conclusion

This section explored the responsible integration of artificial intelligence in mental health care with a focus on accessibility, ethical challenges, emotional depth, and more. By analyzing AI powered tools like Woebot, the report highlights the need for thoughtful implementation to protects users, particularly adolescents and marginalized communities. AI therapy can effectively supplement mental health systems by offering stigma-free support options. Without ethical precautions, these tools risk breaking user privacy, lacking emotional touch, or discouraging future help. AI must be viewed as a supporting tool, not as a replacement for human care.

This research connects to a larger report of AI in health care by discussing the advantages and risks of expanding these tools into mental health care. Just as AI supports diagnostics in broader health care contexts, as shown by Johnson & Johnson's work, mental health applications must prioritize accuracy, empathy, and human monitoring. These tools must continue to develop to uphold ethical standards, set expectations, and offer hybrid options. Developers should focus on adaptive and user informed tools that include live support and more options for help. These tools should also have data storage rules and consent standards. The care teams should advocate for models that have the digital convenience but also human empathy that these tools currently lack. These tools should be designed to adapt to users' emotional states and detect disengagement. Ultimately, if implemented properly, these tools can help close the treatment gap and help users in their journeys. This future depends on continued research, transparent regulation, and a commitment to patients and autonomy in every step.

References

Alfano, L., Malcotti, I., & Ciliberti, R. (2024). Psychotherapy, artificial intelligence and adolescents: ethical aspects. *Journal of Preventive Medicine and Hygiene*, *64*(4), E438–E442. https://doi.org/10.15167/2421-4248/jpmh2023.64.4.3135

Bethune, S. (2022, November 15). *Increased need for mental health care strains capacity*. Apa.org; American Psychological Association. https://www.apa.org/news/press/releases/2022/11/mental-health-care-strains

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, *4*(2). https://doi.org/10.2196/mental.7785

Johnson & Johnson. (2023, September 14). *6 ways Johnson & Johnson is using AI to help advance healthcare*. Content Lab U.S. https://www.jnj.com/innovation/artificial-intelligence-in-healthcare

Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H.-C., Paulus, M. P., Krystal, J. H., & Jeste, D. V. (2021). Artificial Intelligence for Mental Healthcare: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(9), 856–864. https://doi.org/10.1016/j.bpsc.2021.02.001

Terra, J. (2024, February 29). *What Is an AI Chatbot, and How Do They Work? - Caltech*. Caltech. https://pg-p.ctme.caltech.edu/blog/ai-ml/what-is-ai-chatbot-how-do-they-work