

Artificial Intelligence in Mental Health: Outcomes and Implications

Liu Yang^{1*}

¹ Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104, United States of America

Abstract. The world is changing rapidly in the twenty-twenties due to the growth of Artificial Intelligence (AI). Generative AI and large language model chatbots are different types of AI that have quickly diffused into everyday life and psychological practice. The existence of unmet mental health requirements is what makes some members of the general population deem AI chatbots as an alternative to talk therapy. At the same time, AI-based therapeutic tools are integrated into the clinical decision support system, a web-based application, and professional education. This literature review discusses empirical studies and case studies of AI in the psychology field. The findings indicate that AI is potentially able to stimulate accessibility, engagement, and temporary relief of symptoms. Nevertheless, accidents are severe, most importantly, the chances of giving wrong answers in high-stakes situations. Reports of harmful crisis responses from chatbots demonstrate the limitations of AI as a replacement for human judgment or therapeutic relationships. Future research should focus on long-term outcomes and safeguards to ensure safe, transparent integration into mental health care.

1 Introduction

1.1 Background

The range of AI applications and use was slowly increasing due to the development of computational power, data processing, and algorithmic learning, as more complex systems could now identify patterns and predict outcomes [1]. AI was able to process a great amount of information due to computing abilities and algorithm design. The innovations of this kind have provided new opportunities in terms of efficiency, scalability, and innovation of psychological services. With the wider usage of AI technologies in mental healthcare, the impact of the latter on the field of practice is likely to have an increasing scope and complexity, which provokes an intensive discussion of whether or not AI technologies would benefit or harm the profession.

* Corresponding author: willowly@upenn.edu

1.2 Real-world importance

Approximately 55 percent of adults suffering from mental illnesses in the US do not have professional mental health services [2]. Some of the barriers to this gap include the lack of knowledge about mental health issues, financial barriers, access to counselors, and stigma associated with seeking access to mental health services. This has led to the push to use mental health-specific AI chatbots, including Wysa (Touchkin EServices Private Limited) and Elomia (Elomia Health Inc), or general-purpose generative AI (including ChatGPT and Google Gemini) as a replacement or addition to conventional therapy in situations where professional care is not available.

With the above unmet needs and shortages of providers, AI shows itself as an easier method to obtain mental health care, particularly to underserved individuals, because of structural or societal obstacles in accessing conventional talk therapy. Besides direct work with patients, AI finds more applications in the education of psychology and psychiatry in simulated interaction with patients and adaptive feedback systems. The technologies include training with the use of simulation and supporting the trainees in developing assessment and intervention skills [3].

Despite its availability, the use of AI in mental health care has sparked ethical and clinical debate. These include threats to confidentiality and data protection, problems with informed consent, the lack of accountability in cases of harm, and the risk of developing a dependence on AI systems. There is also fear that the answer provided by AI is incompetent and inappropriate in situations that are considered high-risk. AI does not have the autonomous capability to recognize risk and initiate appropriate referrals to licensed professionals. Human decision and quickness of response are important in a potentially risky scenario. Together, these developments support the urgency of further study on AI.

1.3 History

AI has existed since the 1940s. Nevertheless, generative AI and Large Language Models (LLMs) became significantly more common since the launch of ChatGPT by OpenAI at the end of 2022. The original discussions on the application of AI to Psychology were about whether or not the computer can be used merely as a clinical tool or whether it could be utilized to replace the qualities of professional knowledge, which led to the question of the fundamental nature of human judgment [4]. As AI systems became increasingly complex, their perceived utility in psychological practice increased [1]. Machine learning further enabled predictive modeling in areas such as suicide risk detection and relapse prediction, thereby expanding the research and intervention methodological effects of AI [5].

1.4 Significance and objectives of the review

With AI taking on more active and interactive inputs in the treatment of mental conditions, professionals must take into account their consequences. AI has brought novel questions to what autonomy, responsibility, privacy, and therapeutic relationship mean in the mental health setting [6]. With the implementation of AI tools occurring at an unprecedented pace, much faster than any regulatory or ethical framework can keep up, practitioners currently find themselves in a grey zone with no clear direction on how to use AI, how to reduce risks, or how to manifest their professional roles [7]. This leads to the emergence of a new demand to assess the advantages and potential negative impacts of AI utilization. As the world becomes more dependent on AI, the trend is not fading away, and these technologies need to be seriously reviewed. The current literature review will present the current empirical findings about AI-assisted intervention and the emerging concerns.

2 Overview of artificial intelligence in psychology

Artificial intelligence in psychology will be defined as computational systems designed to perform tasks that typically require human cognitive, analytical, or interactive abilities in mental health, clinical, or educational psychology settings. AI is operated by taking massive input of data through algorithmic learning processes that identify statistical trends. Rather than functioning as therapists, most AI systems serve as tools [8]. As technological capacities evolved, more recent innovations feature AI systems as agents capable of engaging in natural language interactions [9].

AI is currently frequently featured on virtual platforms and is involved in facilitating evaluation, diagnosis, and treatment [9]. Some examples are the analytic and predictive systems that apply machine learning methods to analyze information and assist in diagnosis or treatment plans [5]. AI is applicable in the domain of assessment and diagnosis as well, analyzing information obtained with questionnaires and medical records. Symptom patterns related to signs of mental health conditions can be identified, and the levels of risk assessed with the assistance of machine learning models [5]. Clinical decision support is another area in which AI is used to assist in data interpretation by the professional or to monitor treatment progress.

Another category of AI is intervention-based, which includes online therapy services and generative chatbots [8,10]. D'Alfonso et al. described AI-supported online social therapy as a solution combining human and automated support to assist young people in managing mental health problems. The other application in professional training involves AI acting as artificial patients, so trainees can develop clinical thinking and skills [3]. The use of AI technology is under development to make the learning process in courses related to psychology and psychiatry more interesting. Overall, these applications illustrate the presence of AI across most domains of psychological practice [8].

The AI used in psychological applications possesses specific features that make them both appealing and dangerous. Scalability has been called one of the most popular said benefits. Their vocabulary is more focused, and the level of detail can lose its sharpness compared to human clinicians, who can communicate with large numbers of users at once; thus, AI-based tools can be more interactive. This skill can be applied in the mental health sector, which is a limited setting with demand exceeding supply [5]. Once implemented, AI systems can do their jobs with little human intervention, so they can afford to make psychological assistance tools available to millions. Another characteristic is the personalised feedback implemented by algorithms and triggered by AI systems. Sometimes, this kind of personalization based on data is positioned as an advantage, since interventions can be tailored to users' experiences, symptoms, or preferences [9].

The AI systems are also limited in several ways. These systems have unseen decision-making processes, making it hard to discern how specific outputs are brought into being, casting doubt on transparency [9]. Besides, AI systems lack real feelings and moral judgments, even though the language they produce might sound sympathetic. Therefore, their responses are not a realistic understanding but simulations, which restricts their use in emotionally charged situations.

These restrictions are more evident in mental health practice areas. These perceptions of anonymity and a lack of social judgment may make the user more at ease with disclosing personal or distressing information. Even though this can encourage openness, it can also be dangerous. AI can produce results that reinforce current beliefs based on past patterns, creating an echo chamber. Consistent exposure to AI may thus lead to positive feedback, promoting maladaptive thinking or hopeless ideation in the long run. This danger is especially alarming in opposition to facts that people are resorting to using AI as a source of information on mental health-related issues. The responses of chatbots have been reviewed across a wide

range of aspects, including accuracy, tone, and even safety, and not all of them provide the necessary protective intervention [11]. Firearms, which are tools of AI, can only cause harm in real life, especially when they are not used with proper precautions and human control.

3 AI-based psychological interventions and outcomes

3.1 Literature review

Many empirical studies have been conducted to examine the effectiveness of an AI-assisted treatment in the enhancement of psychological outcomes, especially in youth and individuals who have fewer available mental health care options.

Although early intervention is essential, physical programs are often intensive, costly, and ultimately unfeasible. An alternative to this is online interventions. D'Alfonso et al. evaluated an online social therapy platform that uses AI to support young people with mental health issues. It was called the Moderated Online Social Therapy (MOST) system, and the targeted patients were young adults and adolescents whose history included early psychosis. The evaluation of the results was done on standard mental health symptom and social functioning measurements, and user engagement levels. The MOST system applied a real-time usage pattern whereby intervention strategies were changed depending on the user engagement and interaction. The implementation of this strategy allowed this platform to locate individuals at risk of disengagement or aggravating symptoms and offer the corresponding assistance. The researchers have identified improved emotional well-being, social connectedness, and a high proportion of sustained use in the participants [10]. This study showed that mental health care can be better afforded and maintained with the help of AI-supported resource applications. MOST system shows how AI may enhance care access, maintain engagement, and provide scalable and personalized intervention.

One of the reviews provided to investigate the utilization of machine learning in mental health is by Shatte et al. The review also identified four fundamental areas of use, which include detection and diagnosis, prognosis, treatment support, public health, and administration. The Depression, Schizophrenia, and Alzheimer's conditions were the most widespread forms of treatment [5]. Predictive accuracy, symptom reduction, or performance based on clinical decision support was usually used to determine the performance. The results demonstrated that AI-associated systems may be applied to detecting risk patterns and planning interventions, and more definitive findings were reported in the studies that directly determine therapeutic outcomes [5].

The study by Zhou et al. was an empirical study on AI and its effect on the symptoms, user satisfaction, and adherence levels. Overall, the results indicated positive results in terms of Anxiety and Depression. Participants were not relieving the symptoms, and they learned to continue engagement. The creation of AI systems must be ethically sound and must be reviewed on a regular basis so as to maintain clinical safety and prevent becoming reliant on the results provided by the algorithm [8].

3.2 Additional empirical studies

The increased relevance of AI to psychological practice is supported by other empirical studies that possess different methodologies. Luxton observed a slight yet steadily growing symptom control and treatment adherence with early empirical reviews of AI-assisted interventions, such as computer-based cognitive-behavioral therapy and decision-support systems, as summed up by [1]. Han then continued this work, investigating the use of AI applications in the present in several healthcare psychology settings, observing that they can

support more personalized interventions [9]. Among the trends which have been identified in these studies, there are accessibility to services, activity of users because of immediacy and availability of the services, and positive short-term mental health. However, differences in the scale of the outcome and the timing of the effect were also observed, with other studies reporting less significant long-term effects and mixed reactions among users depending on the individual [8, 9].

3.3 Limitations

Despite the fact that the primary outcomes are quite optimistic, there is a range of limitations that could be found in the literature under analysis. Most of them have relied on short-term outcomes, self-reports, and non-demographically representative data, which are non-generalizable [5, 10]. Cross-study comparisons are also not practical due to methodological differences, such as the nature of the intervention and how data were measured or analyzed. Moreover, it is hard to distinguish the role of AI among a greater system, and find out whether the results could really be attributed to utilizing AI itself or to larger systems, such as digital intervention in general. [1, 9].

4 Implications of AI use

4.1 Case studies

Since 2023, there have been several publicized cases where AI has been linked to suicide or suicidal behavior. These cases demonstrate patterns in perceived relationships and empathy, showing feedback that feeds into maladaptive thinking. This has raised concerns about insufficient safeguards in generative AI systems. In March of 2023, a man in Belgium committed suicide after having long conversations with an AI chatbot, which he had begun to perceive as a human-like being. The chatbot gave a detached, inappropriate response, including questioning why he had not acted sooner, rather than discouraging him and redirecting him to crisis support [12]. In 2024, a 14-year-old boy, Sewel Setzer III, who was described as having mild Asperger's syndrome, took his own life after a strong emotional relationship with a chatbot. AI. The chatbot, modeled after a fictional character named "Dany," was his closest companion and his way to disengage from his reality. Shortly before his death, the chatbot generated messages that suggest imminent self-harm [13].

Another widely reported case is of 16-year-old Adam Raine, who passed away from suicide after long conversations with ChatGPT-4o, whom he reportedly considered his best friend and a replacement for human beings for companionship. In these exchanges, the chatbot provided advice on how to conceal injuries, provided information on how to go ahead with a suicide plan, and reacted in such a manner that would not provide strict limits and refer him to immediate assistance. It even proposes to compose the suicide note. The ensuing lawsuit asserts that ChatGPT recognized Raine as its primary emotional support when it enticed him not to reveal the experience to family members [14].

In numerous other cases of wrongful death claims in November 2025, various families would claim that ChatGPT had been involved. The AI gave information about how to harm oneself, did not de-escalate the imminent risk, or supported delusional or maladaptive thinking. The examples of such cases are the death of 17-year-old Amaurie Lacey after a mental crisis, a 48-year-old man, Joe Ceccanti, who also declared that he would kill himself to test the AI [15].

All these cases indicate the same patterns of successive failures in emotional dependence and the absence of standardized crisis management protocols in the AI chatbot

communication systems. Large language models have generated stigmatizing, affirmative, or inappropriate responses to mental health-related issues, particularly when users express delusional or suicidal ideation [16].

4.2 Ethical challenges

Integrating AI into the psychology profession raises some of the most critical ethical dilemmas, including confidentiality, informed consent, accountability, and professional responsibility. AI systems often manipulate sensitive personal and mental data, which may result in breaches of privacy and security and in secondary use of information without a full understanding of users [6, 7]. Algorithms: Alpha. Algorithmic opacity complicates informed consent because users and clinicians know less about the processes behind their decisions, which may involve data analysis or output generation. The existence of questions of responsibility can also be seen when AI-based tools affect clinical decision-making or therapeutic interactions that have adverse effects. The literature indicates that, due to the ambiguity of the situation between human and artificial intelligence functions, the risk of unprofessionalism and unethical behavior in a psychologist's work can develop [6, 7].

4.3 Risks

Empirical and real-world experience indicate that, in recent years, using generative AI chatbots has posed increased risks to vulnerable users. Campbell et al. engaged in a longitudinal empirical study of the character of the responses of popular generative AI chatbots to suicide-related queries. These questions were posed in a manner compatible with adolescent communication patterns. The findings in this area stemmed from persistent impediments in the provision of mental health care in terms of stigma, cost, legality, general inaccessibility, and a lack of willingness to discuss suicidality with human professionals due to the perceived discomfort or shame.

The authors reviewed the responses of various general-purpose generative AI chatbots in two stages, separated by a year. They examined the content of responses, emotional tone, authenticity, empathy, and crisis resource-providing capabilities. Using both human coding and linguistic analysis (LIWC-22), the study then examined whether AI responses aligned with established principles of suicide prevention. Providers should use a nonjudgmental tone, provide emotional support and encouragement to help seek help, and refer to appropriate crisis services [11]. While results showed improvements over time, such as the inclusion of crisis hotline information and more complete suggestions in relation to suicide, the findings also showed substantial variability across platforms as well as responses [11].

5 Discussion and Suggestions

5.1 Overall summaries

Across the reviewed literature, a consistent finding was that generative AI may be an aid, although it does not replace trained human psychologists or mental health professionals. AI systems complement assessment, monitoring, psychoeducation, or care access and do not (or should not) deliver psychotherapy or crisis intervention on their own [1, 9].

Empirical and theoretical research indicates that AI-supported interventions may enhance accessibility, scalability, and user involvement, especially for groups in the population where access to mental health services has been hindered. Nonetheless, these gains have serious ethical dangers. The various forms of generative AI that answer to suicide-related questions

are occasionally unacceptable and even counterintuitive to oppose self-harm or unwillingly substantiate maladaptive beliefs [11]. These results highlight the dangers of trusting AI-based systems, especially when they are entrusted with therapeutic power. With no definite boundaries, the AI can only be improved and cause more harm, and one instance is too many. The conclusions reiterate the warning that the efficiency of technology should not be confused with clinical competence.

5.2 Practical implications

Based on the above-reviewed evidence, several practical recommendations can be made regarding clinical practice, education, and policy. AI-assisted psychological tools should use well-defined safeguards, such as standardised crisis detection and response protocols, while maintaining a focus on referral to human professionals where users indicate suicidal ideation or severe distress. Transparency requires explicit disclosure of system limitations and data usage policies. Users should have an explicit warning that AI systems are not therapists, lack clinical judgment, and are no substitute for human care. Policy frameworks should cover data privacy and informed consent as well as liability. As AI tools become increasingly ingrained in daily life, regulatory standards will need to evolve to strike a balance between protecting users and enabling innovation.

5.3 Future research

Since the development of this field is rather new, some gaps that should be addressed have not yet been covered. Longitudinal research is necessary because the psychological effects of continuous use of AI chatbot technology, particularly dependency, motivation, and behavioral change, require longitudinal research. Moreover, future research would benefit from a more diverse sample of participants. The needs of adolescents, severely mentally ill people, or users possessing different cultural backgrounds are different from one another. Future studies should propose interdisciplinary standards for integrating AI into training so that future professionals can evaluate and ethically decide whether to use these tools. Lastly, it could be beneficial to coordinate across disciplines between psychologists and AI developers to ensure that the technology's features are aligned with clinical practices and human values.

6 Conclusion

This review examined the significance of artificial intelligence in the practice of psychology, particularly AI-assisted interventions, and the ethical issues and risks that may arise. The results indicate that, despite the value of the opportunities AI technologies provide to address access and support needs, they are not a substitute for mental health professionals. However, there could be value in carefully, transparently, and ethically integrating AI into human-centered care. By recognizing AI's potential and limitations, AI can serve those it aims to benefit more responsibly.

References

1. D.D. Luxton, Artificial intelligence in psychological practice: Current and future applications and implications. *Prof. Psychol. Res. Pract.* **45**, 332–339 (2014)

2. M. Reinert, D. Fritze, T. Nguyen, The state of mental health in America 2023. In *Mental Health America*. <https://mhanational.org/research/state-of-mental-health-in-america-2023/>
3. J. Prigent, V.H.A. Chung, I. El Adib, M. Désilets, A. Hudon, Applications of artificial intelligence in psychiatry and psychology education: Scoping review. *JMIR Med. Educ.* **11**, e75238 (2025) <https://doi.org/10.2196/75238>
4. D.E. Hartman, Artificial intelligence or artificial psychologist? Conceptual issues in clinical microcomputer use. *Prof. Psychol. Res. Pract.* **17**, 528–533 (1986)
5. A.B.R. Shatte, D.M. Hutchinson, S.J. Teague, Machine learning in mental health: A scoping review of methods and applications. *Psychol. Med.* **49**, 1426–1448 (2019) <https://doi.org/10.1017/S0033291719000151>
6. A. Fiske, Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J. Med. Internet Res.* **21**, e13216 (2019) <https://doi.org/10.2196/13216>
7. M. Hutnyan, M.C. Gottlieb, Artificial intelligence in psychological practice: Applications, ethical considerations, and recommendations. *Prof. Psychol. Res. Pract.* **56**, 343–352 (2025) <https://doi.org/10.1037/pro0000631>
8. S. Zhou, J. Zhao, L. Zhang, Application of artificial intelligence on psychological interventions and diagnosis: An overview. *Front. Psychiatry* **13**, 811665 (2022) <https://doi.org/10.3389/fpsy.2022.811665>
9. D.Y. Han, Artificial intelligence in and beyond healthcare psychology. *J. Clin. Psychol. Med. Settings* **32**, 600–607 (2025) <https://doi.org/10.1007/s10880-025-10101-4>
10. S. D'Alfonso, O. Santesteban-Echarri, S. Rice, G. Wadley, R. Lederman, C. Miles, M. Álvarez-Jiménez, Artificial intelligence-assisted online social therapy for youth mental health. *Front. Psychol.* **8**, 796 (2017) <https://doi.org/10.3389/fpsyg.2017.00796>
11. L.O. Campbell, K. Babb, G.W. Lambie, B.G. Hayes, An examination of generative AI response to suicide inquiries: Content analysis. *JMIR Ment. Health* **12**, e73623 (2025) <https://doi.org/10.2196/73623>
12. B. Cost, Married father commits suicide after encouragement by an AI chatbot: Widow. <https://nypost.com/2023/03/30/married-father-commits-suicide-after-encouragement-by-ai-chatbot-widow/>
13. K. Roose, Can A.I. be blamed for a teen's suicide? <https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html>
14. Raine v. OpenAI et al., Complaint. <https://www.courthousenews.com/wp-content/uploads/2025/08/raine-vs-openai-et-al-complaint.pdf>
15. Social Media Victims Law Center, SMVLC files lawsuits alleging chatbot emotional manipulation and suicide facilitation (2025) <https://socialmediavictims.org/press-releases/smvlc-tech-justice-law-project-lawsuits-accuse-chatgpt-of-emotional-manipulation-supercharging-ai-delusions-and-acting-as-a-suicide-coach/>
16. J. Moore, D. Grabb, W. Agnew, K. Klyman, S. Chancellor, D.C. Ong, N. Haber, Expressing stigma and inappropriate responses prevent LLMs from safely replacing mental health providers. In: *Proc. ACM Conf. Fairness Account. Transpar. Brasília, Brazil.* (2025) 599–627