

Mental Health Services in the Age of Artificial Intelligence:

Embrace the Benefits,
Beware of Risks



Table of Contents

Introduction..... 1

Artificial Intelligence—How Machines Learn..... 2

Uses of Artificial Intelligence in Mental Health Care 2

Oversight of AI Mental Health Support Tools.....4

Conclusion 5

References.....6





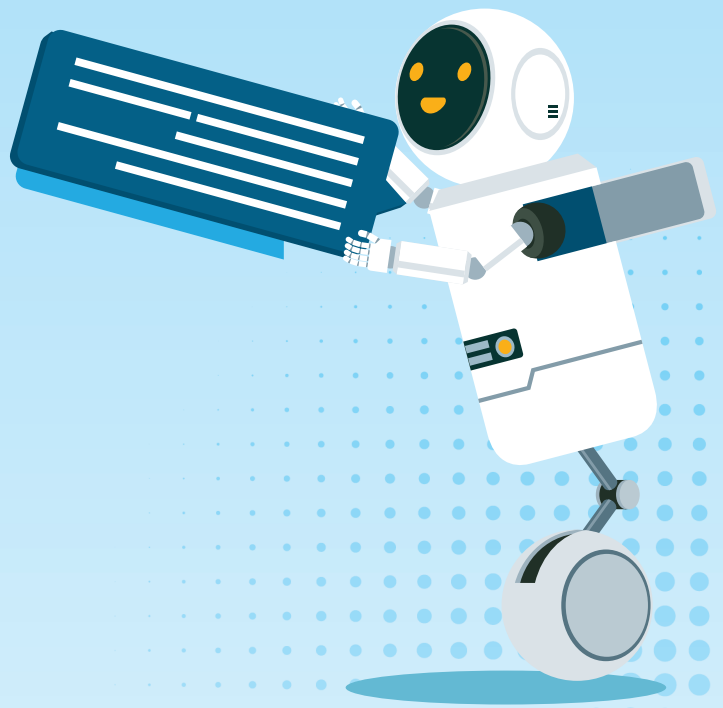
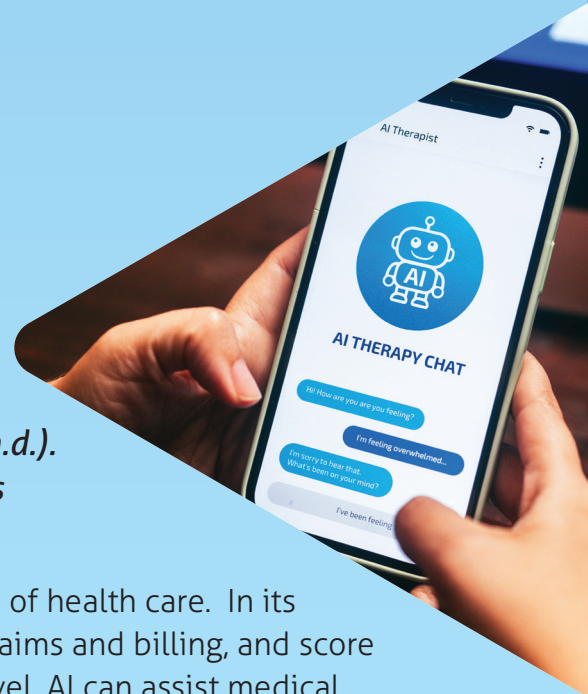
Introduction

Artificial Intelligence (AI) refers to technology that allows computers to replicate human learning, reasoning, decision-making, and communication (Stryker & Kavlakoglu, n.d.). Like humans, AI draws on past experiences and applies lessons learned to enhance its skills and performance over time.

AI technology is rapidly opening new possibilities for the provision of health care. In its simplest form, AI can schedule appointments, manage insurance claims and billing, and score clinical questionnaires (Abrams, 2025). At a more sophisticated level, AI can assist medical professionals in diagnosing illnesses, documenting medical notes, monitoring symptoms, and conversing with clients (Ellis, 2024).

Currently, therapeutic AI mental health services are not federally regulated and criteria for clinical efficacy have not been established by the Food and Drug Administration (Walters, 2025). Nevertheless, AI mental health services appeal to many consumers because they are affordable (and sometimes free) and accessible at any time, and remove barriers pertaining to stigma, shame, and the shortage of mental health providers.

As with all new innovations, both predictable and unanticipated benefits and risks are emerging from AI activities (Gutierrez et al., 2024). This paper discusses the status of identified strengths and weaknesses of AI support in mental health service environments.



Artificial Intelligence—How Machines Learn



AI learning begins with a model (or algorithm) that is designed to make decisions and predictions about selected issues by analyzing and identifying patterns from large amounts of data. As new data sets are introduced into the model, AI performance is refined without the need for additional programming. This process is referred to as machine learning (Stryker & Kavlakoglu, n.d.).

Another component of AI is Natural Language Processing, in which computers learn to recognize, understand, interpret, and respond to human language (Stryker & Holdsworth, n.d.). AI technologies that communicate with humans in this way are commonly known as AI Companions. When individuals engage with AI Companions, the companion remembers and

integrates information from previous conversations into ongoing discussions; and over time, the companion can respond in a way that feels personal and relevant.

The choice of data that perpetuates machine learning impacts the objectivity and quality of AI (Agarwal et al., 2022). In the social science field, if data are collected from limited demographic groups, outputs may not adequately represent the diversity of life experiences across different populations. If data are collected from web sites that discriminate against those with mental health conditions, unintended biases may arise and lead to errors in AI's decision making. Once this information is incorporated into AI training, it is difficult to minimize the damage. Currently, because underlying sources of data inputs and machine learning processes are infrequently disclosed, users are unable to assess the objectivity of AI Companion content (Lee et al., 2021).

Uses of Artificial Intelligence in Mental Health Care

ADMINISTRATIVE TASKS

Mental health clinicians have embraced AI to streamline administrative tasks that are cumbersome and time-consuming. AI can facilitate appointment scheduling, communicate appointment reminders, and ask clients to confirm planned appointments (American Psychological Association, 2024). AI has allowed clinicians to automate intake forms and organize progress notes from multiple client sessions to reveal progress trends and provide insights that may otherwise go undetected. The use of AI for office management purposes improves efficiencies without compromising the client-clinician relationship or value of direct mental health services.

MENTAL HEALTH ASSESSMENT AND DIAGNOSTIC SUPPORT

A strength of AI is its ability to quickly organize massive quantities of information, such as results of health screenings and client interviews, medical histories, and previous treatment effectiveness (Stringer, 2025). AI can also interpret this information, generate a working diagnosis, and recommend treatment interventions.

Although artificial intelligence (AI) tools are becoming increasingly sophisticated, their use in generating diagnoses and treatment plans without professional oversight remains inadvisable. Diagnostic evaluation in mental health care is inherently complex and requires nuanced interpretation. In addition to the risk of algorithmic bias (see Artificial Intelligence—How Machines Learn above), current AI systems lack the capacity to accurately interpret body language and communication styles, components that are integral to a comprehensive assessment. The omission or misinterpretation of such elements by AI systems may result in diagnostic inaccuracies and the formulation of ineffective treatment strategies, thereby compromising the quality of care (Zhang & Wang, 2024).

TREATMENT PLAN SUPPORT

Many AI tools have been developed to help clients track symptoms, practice coping skills, and monitor progress towards treatment plan goals. They also provide feedback, including accolades for improvement and



suggestions if improvement is not sufficient.

When a client and professional clinician coordinate to establish a proper diagnosis and treatment plan, the clinician can recommend a preferred tool to deliver psychoeducation and structured exercises to augment work that occurs during therapeutic sessions (American Psychological Association, 2024). Choosing a tool without a treating clinician's input may conflict with treatment plan interventions and cause frustration by the client.

AI COUNSELING AND PSYCHOTHERAPY

AI counseling and psychotherapy (referred to as therapy hereafter), involves the use of AI to establish relationships with clients, learn about their current problems, and provide guidance on how to address problems.

AI therapy is available to all who have access to an internet-connected computer or smartphone. An AI therapist may be contacted at any time of the day and will respond instantaneously. AI therapy is less expensive than face-to-face therapy and reduces stigma and logistical barriers (e.g., time constraints, travel arrangements) (Zhang & Wang, 2024).

But is AI therapy synonymous with face-to-face support by a licensed mental health clinician? Research demonstrates that fundamental differences exist between AI and person-to-person support.

Emotion and Empathy

According to the American Psychological Association's Dictionary of Psychology (2023), empathy

is defined as "recognition and understanding of an individual's feelings, perceptions, and thoughts."

Empathy is a necessary component of an effective therapeutic relationship. While AI can recognize and acknowledge emotional reactions through word patterns, its understanding of emotions is different from humans. AI relies on language recognition and modeling, not genuine feelings or lived experiences (Zhang & Wang, 2024). Consequently, AI cannot replicate the depth of empathy necessary for building therapeutic alliance.

Stigma

A recent study by Stanford University researchers found that AI systems exhibit stigmatizing responses toward individuals with mental health conditions, reflecting biases embedded in training data. Increasing model scale does not significantly reduce these biases (Moore et al., 2025). Such findings underscore the ethical challenges of deploying AI in mental health care without rigorous oversight.

Client Safety

In the same study, researchers assessed whether AI therapist responses appropriately identify and address serious symptoms such as delusions, suicidal ideation, and obsessive-compulsive behaviors. The study found that responses were inappropriate, and dangerous in twenty percent of experimental scenarios. In addition, the study concluded that AI therapists may encourage perpetuation of delusional thinking and inadvertently enable

harm to self/others (Moore et.al. 2025).

Professional Codes of Conduct

Research from Brown University (2025) revealed that AI therapy violates several ethical standards outlined by the American Psychological Association. These violations include creating a false sense of caring and trust, failing to adequately address safety concerns, disregarding individualized treatment planning, and endorsing inaccurate beliefs. Such shortcomings highlight the inability of AI systems to meet professional codes of conduct governing psychological practice.

Trust and Confidence

Some clients may mistakenly believe that AI therapy offers services equivalent to those provided by licensed clinicians. Unlike human therapists who foster autonomy and self-awareness through nuanced interpersonal interactions, AI operates within rigid, pre-programmed parameters. Furthermore, some commercial AI platforms employ strategies that encourage user dependence, contrasting sharply with the therapeutic goal of promoting independence (Khawaja & Belisle-Pipon, 2023).

Believing that AI therapy is interchangeable with human therapy can lead to delayed professional intervention, inadequate treatment, and worsening mental health symptoms. Users may rely on AI for complex issues requiring human expertise—such as trauma, suicidal ideation, or rapid functional decline—resulting in false confidence and compromised outcomes.



Oversight of AI Mental Health Support Tools

The rapid proliferation of AI tools in mental health care has outpaced the development of comprehensive regulatory frameworks, raising significant concerns about safety, efficacy, and ethical compliance. Despite widespread adoption of AI-enabled companions marketed for emotional support, the U.S. Food and Drug Administration (FDA) has not approved any AI-based systems for diagnosing, treating, or curing mental health disorders (Walters, 2025). This lack of approval stems from the absence of clinical trials necessary to establish safety and effectiveness, a prerequisite for FDA clearance of medical devices. While the FDA has cleared many AI-enabled devices across healthcare domains, none currently pertain to mental health, highlighting the complexity of regulating technologies that mimic therapeutic interactions.

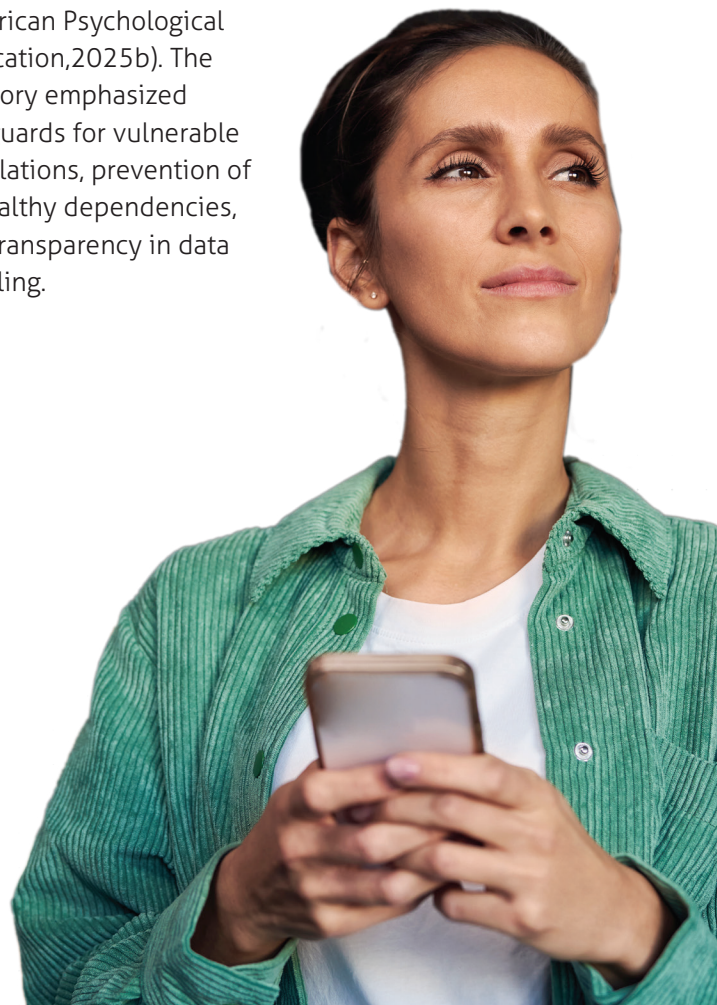
There are no federal laws governing use of AI Companions, but 20 bills were voted into law in 11 states pertaining to licensure requirements, safety protocols, confidentiality and transparency, and data privacy (Shumate et al., 2025).

On November 6, 2025, the FDA's Digital Health Advisory Committee (DHAC) convened to evaluate AI-enabled mental health devices. Key concerns included algorithm bias, privacy vulnerabilities, and the inability of AI systems to reliably manage crises. The FDA signaled its intent to develop regulatory criteria addressing transparency, advertising practices, and risk mitigation strategies for AI mental health tools (FDA DHAC Executive Summary, 2025). Expected time frames for establishing regulatory criteria were not articulated.

Professional organizations have issued position statements and advisories to guide ethical integration of AI in mental health care. The American Psychiatric Association (APA) released its Position Statement on the Role of Augmented Intelligence in Clinical Practice and Research (2024), affirming that AI should function solely as an adjunct to treatment rather than a replacement for clinical services by a licensed provider. The statement called for culturally and linguistically appropriate patient

education, transparency in AI-driven decision-making, and strict adherence to privacy standards.

Similarly, the American Psychological Association (APA) has taken an active role in promoting safeguards. In early 2025, the APA met with federal regulators to address concerns about virtual therapists falsely claiming licensure and the associated risks of inaccurate diagnoses, inappropriate treatment, and exploitation of minors (American Psychological Association, 2025a). The APA urged the Federal Trade Commission (FTC) to mandate ethical standards for AI mental health tools, including prohibitions on deceptive marketing and requirements for clinician involvement in development. In November 2025, the APA published its Health Advisory on the Use of Generative AI Chatbots and Wellness Applications for Mental Health, warning against reliance on AI as a substitute for professional care and outlining responsibilities for developers, researchers, consumers, and regulators (American Psychological Association, 2025b). The advisory emphasized safeguards for vulnerable populations, prevention of unhealthy dependencies, and transparency in data handling.





*Used responsibly,
AI can enhance
mental health services.*

Conclusion

Artificial intelligence (AI) has gained traction in mental health care, offering tools that promise convenience, accessibility, and improvement in well-being. The benefits and risks of AI mental health support continue to emerge and are under evaluation by federal departments and professional organizations. Until regulatory boards and legal entities formalize guidelines, consumers should be cautious when opportunities to use AI mental health tools are presented.

Consumers should be able to answer “yes” to all the questions below to confirm the potential value and safety of AI services under consideration.

- ☒ **YES:** I am connected to a mental health clinician or other professional provider to manage my mental health needs.
- ☒ **YES:** I have partnered with my mental health clinician or other professional provider managing my mental health needs to identify AI support that:
- ☒ **YES:** Is evidence based (evidence of effectiveness and how the tool was tested are clearly stated).
- ☒ **YES:** Has been created to directly address issues I am working with my mental health professional/provider.
- ☒ **YES:** Provides transparent information on potential biases.
- ☒ **YES:** Provides means of maintaining confidentiality and protecting private health information.
- ☒ **YES:** Does not promise miracles (clearly discloses potential benefits and limitations).



References

- Abrams, Z. (2025 June 26). Artificial intelligence is reshaping how psychologists work. American Psychological Association Services, Inc. <https://www.apaservices.org/practice/news/artificial-intelligence-psychologists-work>
- Agarwal, R., Bjarnodottir, M., Rhue, L. Dugas, M., Crowley, K., Clark, J. & Gao, G. (2023, March 1). Addressing algorithmic bias and the perpetuation of health inequities: An AI bias aware framework. Science Direct. Health Policy and Technology 12(1). <https://doi.org/10.1016/j.hlpt.2022.100702>
- American Psychiatric Association (2024). Position Statement on the Role of Augmented Intelligence in Clinical Practice and Research. <https://www.psychiatry.org/getattachment/a05f1fa4-2016-422c-bc53-5960c47890bb/Position-Statement-Role-of-AI.pdf>
- American Psychological Association (2023). Empathy in Dictionary of Psychology. <https://dictionary.apa.org/empathy>
- American Psychological Association (2024, November 21). Artificial intelligence in mental health care. American Psychological Association. <https://www.apa.org/practice/artificial-intelligence-mental-health-care>
- American Psychological Association Services (2025a). Urging the federal trade commission to take action on unregulated AI. American Psychological Association. <https://www.apaservices.org/advocacy/news/federal-trade-commission-unregulated-ai>
- American Psychological Association (2025b). APA Health Advisory on the Use of Generative AI Chatbots and Wellness Applications for Mental Health. <https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-chatbots-wellness-apps-mental-health.pdf>
- Brown University. Science and Technology (2025, October 21). New study: AI chatbots systematically violate mental health ethics standards. <https://www.brown.edu/news/2025-10-21/ai-mental-health-ethics>
- Ellis, L. (2024, August 30) The benefits of the latest AI technologies for patients and clinicians. Harvard Medical School. Professional, Corporate, and Continuing Education. <https://learn.hms.harvard.edu/insights/all-insights/benefits-latest-ai-technologies-patients-and-clinicians>
- Gutierrez, G., Stephenson, C., Eadie, J., Asadpour, K., & Alavi, N. (2024). Examining the role of AI technology in online mental healthcare: opportunities, challenges, and implications, a mixed-methods review. Frontiers in psychiatry, 15, 1356773. <https://doi.org/10.3389/fpsyt.2024.1356773>
- Khawaja, Z., & Bélisle-Pipon, J. C. (2023). Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. Frontiers in digital health, 5, 1278186. <https://doi.org/10.3389/fdgth.2023.1278186>
- 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 Health Care: 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>
- Lucas, D. (2025 September 30.). AI mental health tools face mounting regulatory and legal pressure. Gardner FDA Law Firm. J.D.Supra. com. <https://www.jdsupra.com/legalnews/ai-mental-health-tools-face-mounting-3647334/#:~:text=The%20emergence%20of%20AI%20enabled,standards%2C%20and%20material%20business%20risk.>
- Moore, J., Grabb, Declan, Agnew, W., Klyman, K., Chancellor, S., Ong, D.C., & Haber, N. Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers (23 June, 2025. FAccT '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. 599-627. Association for Computing Machinery. <https://dl.acm.org/doi/full/10.1145/3715275.3732039>
- Shumate, J. N., Rozenblit, E., Flathers, M., Larrauri, C. A., Hau, C., Xia, W., Torous, E. N., & Torous, J. (2025). Governing AI in Mental Health: 50-State Legislative Review. JMIR mental health, 12, e80739. <https://doi.org/10.2196/80739>
- Stringer, H. (2025, January 1). Technology is reshaping practice to expand psychology's reach. American Psychological Association. Monitor on Psychology. <https://www.apa.org/monitor/2025/01/trends-technology-shaping-practice>
- Stryker, C & Holdsworth, J. (n.d.). What is NLP (natural language processing)? IBM Think. <https://www.ibm.com/think/topics/natural-language-processing>
- Stryker, C. & Kavlakoglu, E. (n.d.) What is artificial intelligence (AI)? IBM Think. <https://www.ibm.com/think/topics/artificial-intelligence>
- U.S. Food & Drug Administration. (2025, November 06). Digital Health Advisory Committee (DHAC). Generative artificial intelligence-Enabled digital mental health medical devices. <https://www.fda.gov/media/189391/download>
- Walters, J. (2025, November 13). FDA committee meets on generative AI digital mental health devices. Psychiatric Times. <https://www.psychiatrictimes.com/view/fda-committee-meets-on-generative-ai-digital-mental-health-devices>
- Zhang, Z., & Wang, J. (2024, October 31). Can AI replace psychotherapists? Exploring the future of mental health care. Frontiers in psychiatry, 15, 1444382. <https://doi.org/10.3389/fpsyt.2024.1444382>



Magellan
FEDERAL

