



International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	LARGE LANGUAGE MODEL-BASED CHATBOTS IN MENTAL HEALTH SUPPORT: A NARRATIVE REVIEW OF OPPORTUNITIES, RISKS, AND ETHICAL CHALLENGES
----------------------	--

DOI	https://doi.org/10.31435/ijitss.3(51).2026.6235
------------	---

RECEIVED	16 July 2026
-----------------	--------------

ACCEPTED	23 September 2026
-----------------	-------------------

PUBLISHED	30 September 2026
------------------	-------------------

LICENSE



The article is licensed under a **Creative Commons Attribution 4.0 International License**.

© The author(s) 2026.

This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

LARGE LANGUAGE MODEL–BASED CHATBOTS IN MENTAL HEALTH SUPPORT: A NARRATIVE REVIEW OF OPPORTUNITIES, RISKS, AND ETHICAL CHALLENGES

Monika Łapot (Corresponding Author, Email: monika.lapot98@gmail.com)

COPERNICUS PL Sp. z o.o., St. Adalbert Hospital, Jana Pawła II Avenue 50, 80-462 Gdańsk, Poland
ORCID ID: 0009-0004-6554-9683

Krzysztof Hober

Healthcare Center in Olesno, Olesno, Poland
ORCID ID: 0009-0005-9280-3664

Dominika Krupnik

Clinical Hospital No. 1 in Zabrze, Zabrze, Poland
ORCID ID: 0009-0003-7560-4546

Wiktoria Skowron

John Paul II County Hospital in Wadowice, Wadowice, Poland
ORCID ID: 0009-0004-1674-6082

Mateusz Skowron

Primary Healthcare Center in Kęty, Kęty, Poland
ORCID ID: 0009-0002-1944-1415

Wiktoria Bzdak

Medical University of Białystok, Jana Kilinskiego 1, 15-089 Białystok, Poland
ORCID ID: 0009-0002-7183-4769

Patrycja Szewińska

Friedrich Lange County Hospital in Nowe Miasto Lubawskie, Mickiewicza 10, 13-300 Nowe Miasto Lubawskie, Poland
ORCID ID: 0009-0007-2145-3358

Joanna Górak

COPERNICUS PL Sp. z o.o., St. Adalbert Hospital, Jana Pawła II Avenue 50, 80-462 Gdańsk, Poland
ORCID ID: 0009-0006-5536-8269

Wojciech Neumann

Non-public Healthcare Center Euromedica, Grudziądz, Poland
ORCID ID: 0009-0009-9483-0868

ABSTRACT

Background: Mental disorders represent a major global public health challenge. Mental health care is affected by limited access to services, shortages of mental health professionals, stigma, and growing healthcare costs. In response to these challenges, artificial intelligence-based technologies, including chatbots and large language models, are increasingly being explored as tools to support mental health care.

Aim: The aim of this narrative review was to examine the potential benefits, risks, and ethical challenges associated with the use of artificial intelligence chatbots and large language model-based systems in mental health care. Particular attention was given to their clinical relevance, public health implications, and responsible implementation.

Methods: A narrative review of the literature was conducted using PubMed, Scopus, Web of Science, and Google Scholar. Publications written in English and published between January 2013 and March 2026 were considered. Thematic synthesis focused on technological applications, clinical utility, system-level implications, risks, and ethical and regulatory challenges.

Results: Current evidence suggests that AI-based chatbots may support psychoeducation, emotional support, self-management, symptom monitoring, cognitive behavioral therapy-based exercises, and access to low-intensity mental health support. These tools may be particularly useful in settings with limited access to specialists and for individuals facing barriers related to stigma, cost, or availability of services. However, their use is associated with important risks, including inaccurate or harmful responses, privacy concerns, algorithmic bias, limited transparency, unclear accountability, emotional dependence, and insufficient long-term clinical validation.

Conclusion: AI chatbots and large language model-based systems should be considered adjunctive tools rather than replacements for mental health professionals. Their safe and ethical integration into mental health care requires clinical validation, human oversight, transparent governance, privacy protection, bias assessment, crisis protocols, and continuous monitoring.

KEYWORDS

Artificial Intelligence, Mental Health, AI Chatbots, Large Language Models, Digital Psychiatry, Ethics

CITATION

Łapot, M., Hober, K., Krupnik, D., Skowron, W., Skowron, M., Bzdak, W., Szewińska, P., Górak, J., & Neumann, W. (2026). Large language model-based chatbots in mental health support: A narrative review of opportunities, risks, and ethical challenges. *International Journal of Innovative Technologies in Social Science*, 3(51). [https://doi.org/10.31435/ijitss.3\(51\).2026.6235](https://doi.org/10.31435/ijitss.3(51).2026.6235)

COPYRIGHT

© **The author(s) 2026.** This article is published as open access under the **Creative Commons Attribution 4.0 International License (CC BY 4.0)**, allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

Introduction

Mental health is a fundamental component of individual well-being and population health (World Health Organization [WHO], 2021). From a healthcare system perspective, mental disorders represent a significant and growing public health challenge worldwide (Patel et al., 2018; Vigo et al., 2016). Globally, more than one billion people live with mental disorders, with anxiety and depressive disorders being among the most common conditions (WHO, 2022). Data from the Global Burden of Disease Study indicate that the burden of mental disorders is particularly high among adolescents aged 15–19 years. The study also found that women generally have higher age-standardized mean disability-adjusted life years (DALYs) than men (GBD 2023 Mental Disorder Collaborators, 2026). Mental disorders affect both psychological well-being (Centers for Disease Control and Prevention, 2025) and physical health (Prince et al., 2007). Mental disorders also generate a substantial economic burden. Globally, their costs reach approximately USD 1.9 trillion (Arias et al., 2022). OECD reports indicate significant differences between countries in mental health expenditure (OECD, 2021). Western European countries allocate a substantially higher share of healthcare spending to mental health services than Poland (OECD, 2021). The prevalence of mental health problems in Poland may be underestimated due to lower awareness, stigma, and poorer access to services. The report also states that mental health funding in Poland accounted for just over 3% of National Health Fund (NFZ) expenditures. (OECD, 2023) A key challenge in the care of people with mental disorders is limited access to mental health services (Schwartz, 2025). Despite growing global awareness of the importance of mental health and the implementation of relevant policies, access to services remains limited. This is largely due to substantial underfunding, particularly in low- and middle-income countries (LMICs) (Cruz-Gonzalez et al., 2025;

Dehbozorgi et al., 2025). An additional and critical challenge is the persistent shortage of mental health specialists, which significantly limits timely access to care (Patel et al., 2016). Poland has been among the European countries with the lowest number of psychiatrists relative to population size, with approximately nine psychiatrists per 100,000 inhabitants reported for 2017 (Eurostat, 2020). This shortage is also consistent with national health policy documents, including the Polish Map of Health Needs for 2022–2026, which identifies health system challenges and resource needs at the national and regional levels (Minister Zdrowia, 2021). This shortage contributes to longer waiting times for psychiatric consultations, delays in diagnosis and treatment initiation, and, ultimately, poorer patient outcomes. There is also a lack of well-integrated and sustainable care models that ensure continuity of support for patients. In this context, the development of artificial intelligence-based technologies offers new opportunities to improve psychiatric care. AI may support earlier detection of mental disorders, better personalization of treatment, long-term monitoring of patients' mental health status, and greater patient engagement in the therapeutic process (Cruz-Gonzalez et al., 2025; Dehbozorgi et al., 2025). The persistent stigma associated with mental health care remains a significant barrier to accessing treatment worldwide (Thorncroft, 2017). Research conducted in Poland indicates that negative attitudes toward people with mental disorders and reluctance to seek psychiatric treatment are still present in Polish society (Burdzik, 2023). Due to stigma, shortages of specialists, and underfunding of services, a substantial proportion of people with mental disorders do not receive appropriate care (Henderson et al., 2013). The COVID-19 pandemic further exacerbated these problems. A study published in *The Lancet* estimated that the global prevalence of major depressive disorder increased by 27.6%, while anxiety disorders increased by 25.6% during the pandemic, with women and younger people being the most affected groups (Santomauro et al., 2021). These challenges highlight the need to develop interventions that are accessible, scalable, and cost-effective (Santomauro et al., 2021). Consequently, healthcare systems are increasingly seeking alternative models of care delivery, including telepsychiatry (O'Keefe et al., 2019). In response to these challenges, digital psychiatry has become a rapidly evolving field focused on developing technologies that improve the accessibility, efficiency, and scalability of mental health services (Torous et al., 2021; Firth et al., 2017). It encompasses a range of technologies, including telepsychiatry, mobile health applications, and artificial intelligence (AI)-based tools designed to support diagnosis, monitoring, and treatment (Torous et al., 2021). Advances in AI have accelerated the development of chatbots for mental health support (Gaffney et al., 2019; Boucher et al., 2021). These tools may provide access to support at any time, facilitate self-monitoring of well-being (Oh et al., 2024), reduce perceived stigma, and appeal to users who are accustomed to digital technologies (Bond et al., 2023). Generative AI chatbots, based on large language models (LLMs) such as GPT-based systems, generate responses dynamically rather than relying solely on pre-written messages. By adapting their responses to the user's language, tone, and emotional cues, these systems may facilitate more natural, personalized, and emotionally responsive conversations. Such features may foster greater user engagement (Kumar et al., 2025), support the development of a therapeutic alliance (Siddals et al., 2024), and increase users' sense of being understood (Yin et al., 2024). Chatbots may also support patient engagement and treatment adherence, for example, by assisting users with cognitive behavioral therapy tasks between sessions (Habicht et al., 2025). Beyond structured therapeutic interventions, chatbots designed as virtual companions, such as Replika, have been reported to reduce feelings of loneliness among some users (Liu et al., 2024). Mobile mental health applications can support symptom and behavior monitoring, offering new opportunities for continuous assessment and self-management (Naslund et al., 2016). AI-based approaches have also been used to analyze digital traces and written language from social media platforms to identify markers associated with mental health disorders, including depression (De Choudhury et al., 2013; Coppersmith et al., 2015). Furthermore, speech analysis and emotion recognition systems can analyze voice characteristics, such as pitch, intensity, and speech patterns, to identify emotional states and support the assessment of depression and other mental disorders (Cummins et al., 2015). Virtual reality-based exposure therapy can be used to treat phobias and post-traumatic stress disorder (PTSD) in controlled clinical settings (Maples-Keller et al., 2017). In addition, LLMs are increasingly being explored as tools to support the analysis of clinical questionnaires, psychiatric interviews, clinical narratives, and other forms of mental health assessment (Singhal et al., 2023; Sallam, 2023). Existing reviews suggest that AI-based conversational agents may help reduce mental health symptoms and improve quality of life (Oh et al. 2024; Bhatt et al., 2025; Li et al., 2023; Vaidyam et al., 2019; Zhong et al., 2024; He et al., 2023). However, despite the rapid development of AI applications in mental health care, the literature remains fragmented, particularly with regard to integrating clinical, technological, and public health perspectives into a unified analytical framework (Shatte et al., 2019; Bzdok & Meyer-Lindenberg, 2018). Moreover, although recent reviews have begun to examine

AI chatbots and LLM-based tools in mental health care, further critical synthesis is needed to evaluate their potential benefits, risks, and ethical, regulatory, and societal implications (Floridi et al., 2018; Jobin et al., 2019). Unlike previous reviews that focused primarily on clinical effectiveness or technical performance of AI chatbots, this review integrates clinical, public health, ethical, and regulatory perspectives within a single analytical framework. Therefore, the purpose of this narrative review is to examine the potential benefits, risks, and ethical challenges associated with the use of AI chatbots in mental health care, with particular attention to clinical relevance, public health implications, and responsible implementation.

Methodology

This study adopts a narrative review design aimed at synthesizing existing knowledge on the role of artificial intelligence in mental health care from clinical and public health perspectives.

A literature search was conducted across major scientific databases and academic search engines, including PubMed, Scopus, Web of Science, and Google Scholar. The main search strategy focused on publications written in English and published between January 2013 and March 2026. This time frame was selected to ensure comprehensive coverage of recent developments in artificial intelligence, digital health technologies, and innovations in mental health care. Earlier publications were included selectively when they provided foundational theoretical concepts or important contextual background relevant to human–computer interaction, digital communication, or ethical and social aspects of AI use.

The search terms included combinations of the following keywords: “artificial intelligence”, “digital psychiatry”, “mental health”, “large language models”, “telemedicine”, “AI chatbots”, “ethical challenges”, and “public health”. Additional studies were identified through reference screening of selected articles.

Studies were included if they addressed technological applications of AI in mental health systems, digital mental health interventions, or the ethical and social implications of AI implementation in healthcare. Both empirical studies and high-quality reviews were considered. Non-scientific publications and articles without clear relevance to healthcare applications were excluded.

Given the narrative nature of the review, no meta-analysis or formal quality assessment was performed. The findings were synthesized thematically, with a focus on technological applications, system-level implications, and ethical and societal challenges associated with AI adoption in mental health care.

Description of the Current State of Knowledge

Opportunities and Benefits

ChatGPT has attracted widespread interest because of its ability to generate human-like text and process large amounts of linguistic data (Franco D'Souza et al., 2023). In the field of mental health, many studies on large language models have focused on ChatGPT. These studies have assessed its potential usefulness in various clinical and educational applications, such as supporting diagnostic processes (Franco D'Souza et al., 2023; Wei et al., 2023; Sezgin et al., 2023), treatment planning (Wei et al., 2023; Elyoseph et al., 2024; Berrezueta-Guzman et al., 2024), formulating treatment recommendations (Farhat et al., 2024; Grabb et al., 2023; Perlis et al., 2024), managing patient care (Wei et al., 2023), conducting psychiatric research (Schubert et al., 2023), and providing psychoeducation (Spallek et al., 2023).

Artificial intelligence is increasingly being explored as a tool to support mental health care, particularly in response to the growing demand for services and the persistent shortage of mental health professionals. Many mental health problems, including anxiety disorders, depression, eating disorders, ADHD, and self-harming behaviors, often emerge during adolescence and may persist into adulthood if they are not identified and treated early (WHO, 2022). Recent evidence suggests that AI applications are being developed in many areas of mental health care, including early detection, diagnostic support, symptom monitoring, treatment planning, and prognosis (Sharma et al., 2025).

Several AI-enabled mental health tools have been developed to support psychoeducation, mood tracking, emotional support, CBT-based exercises, mindfulness, and peer support. These applications illustrate the diversity of digital mental health interventions, ranging from structured CBT-oriented chatbots to wellness-focused and companionship-based tools. To illustrate the range of available AI-enabled mental health interventions, Table 1 presents selected tools and their primary functions.

Table 1. Examples of AI-enabled mental health tools and their primary functions.

Tool / application	Main function	Example of use in mental health support
Woebot	CBT-based chatbot	Provides structured cognitive-behavioral therapy exercises, mood check-ins, and psychoeducational support.
Wysa	Emotional support and self-help	Offers conversational emotional support, coping strategies, mindfulness exercises, and stress-management techniques.
Tess	AI-assisted psychological support	Provides automated mental health support for symptoms of anxiety, depression, and stress.
Youper	CBT and mindfulness support	Supports emotional self-monitoring, cognitive reframing, mindfulness, and mood management.
Sanvello	Support for stress, anxiety, and depression	Combines mood tracking, coping tools, relaxation exercises, and self-management strategies.
Moodpath / MindDoc	Mood tracking and psychoeducation	Helps users monitor mood patterns, reflect on symptoms, and access psychoeducational content.
7 Cups	Emotional and peer-based support	Connects users with trained listeners and provides low-threshold emotional support.
Replika	Conversational companionship	Offers empathetic conversation, self-reflection, and perceived social support.

Source: Adapted from Rezaei et al. (2026). Abbreviations: AI, artificial intelligence; CBT, cognitive behavioral therapy.

Research on generative AI chatbots in mental health care has attracted increasing scholarly attention. For example, Zhang et al. (2025) conducted a systematic review and meta-analysis in which they searched 11 databases and identified 5,555 records. They included 26 studies in the systematic review and 14 randomized controlled trials in the meta-analysis. The meta-analysis showed a statistically significant, small-to-moderate effect of generative AI chatbots on reducing negative mental health outcomes ($ES = 0.30$, $P = .047$). The strongest evidence was observed for depressive symptoms, for which the effect reached statistical significance ($ES = 0.49$, $P = .041$). Zhang et al. (2025) also found that socially and emotionally oriented chatbots were more effective than task-oriented chatbots. The authors suggest that social support, empathy, emotional validation, and the feeling of being understood may have a protective effect on mental health. Therefore, when designing AI chatbots for mental health support, informational accuracy alone may be insufficient. Interaction style also appears to be important, particularly when it is empathetic, supportive, and socially oriented.

In another study, Rezaei et al. (2026) cited a review covering 53 studies and 41 different chatbots. The review found that the dominant applications of mental health chatbots were treatment, education, and screening, mainly in the areas of depression and autism.

To provide an overview of the current evidence base, Table 2 summarizes selected key studies evaluating the effectiveness, clinical utility, safety, and limitations of AI chatbots and large language model-based tools in mental health care.

Table 2. Summary of key evidence on AI chatbots and large language models in mental health care.

Domain	Study	Design	AI tool	Key findings	Limitations
Clinical effectiveness	<i>Fitzpatrick et al. (2017)</i>	RCT	Woebot, CBT chatbot	Woebot users showed short-term reduction in depressive symptoms compared with users receiving psychoeducational materials only.	Short intervention period; self-selected sample; mainly mild to moderate symptoms; limited long-term follow-up.
Clinical effectiveness	<i>Abd-Alrazaq et al. (2020)</i>	Systematic review and meta-analysis	Mental health chatbots	Chatbots showed potential benefits for improving symptoms of depression, anxiety, stress, and psychological well-being.	High heterogeneity of interventions and outcomes; limited evidence on long-term effectiveness and safety.
Clinical effectiveness	<i>Zhang et al. (2025)</i>	Systematic review and meta-analysis	Generative AI chatbots	Generative AI chatbots showed small-to-moderate effect on reducing negative mental health outcomes, with the strongest effect for depression.	Limited long-term data; heterogeneity between studies; insufficient evidence in high-risk and severe psychiatric populations.
Digital mental health support	<i>Gaffney et al. (2019)</i>	Mixed-method systematic review	Conversational agents	Increased access to mental health support and treatment assistance	Limited number of high-quality trials; insufficient evidence for severe mental disorders and real-world implementation.
Psychoeducation and self-management	<i>Boucher et al. (2021)</i>	Review	AI chatbots	Support psychoeducation, symptom monitoring, emotional support, and self-management.	Limited clinical validation; concerns about safety, engagement, and integration into healthcare pathways.
Diagnostic and clinical reasoning	<i>Franco D'Souza et al. (2023)</i>	Clinical vignette evaluation	ChatGPT	ChatGPT demonstrated substantial psychiatric knowledge and provided acceptable responses in analyzed psychiatric case vignettes.	Simulated cases rather than real patients; no assessment of real-world clinical outcomes; performance may depend on prompt formulation.
Diagnostic and treatment support	<i>Wei et al. (2023)</i>	Clinical evaluation study	ChatGPT / LLM	The model showed potential usefulness in diagnostic reasoning, treatment planning, and patient-care-related tasks.	Risk of inconsistent responses; limited external validation; should not replace clinical judgment.

Domain	Study	Design	AI tool	Key findings	Limitations
Between-session support	<i>Habicht et al. (2025)</i>	Real-world observational study	Generative AI therapy tool	The tool may improve patient engagement and support therapeutic tasks between sessions.	Observational design; lack of randomization; possible selection bias; limited causal inference.
Empathy and user engagement	<i>Beredo et al. (2022)</i>	Expert evaluation study	VHope, DialoGPT chatbot	Responses were rated by experts as accurate, natural, and empathetic.	Expert-rated responses rather than patient outcomes; limited clinical validation; no long-term follow-up.
Prompt sensitivity and reliability	<i>Grabb et al. (2023)</i>	Prompt engineering case study	Large language model	Small changes in prompt wording may substantially influence model responses and recommendations.	Case-based evaluation; limited generalizability; highlights the need for clinical oversight and careful prompt design.
Safety of treatment	<i>Perlis et al. (2024)</i>	Expert evaluation of AI-generated recommendations	GPT-4 Turbo	The model occasionally generated recommendations judged by experts as less accurate or potentially contraindicated	Risk of clinically inappropriate advice; LLMs should not be used autonomously for treatment planning.
Privacy, bias, and governance	<i>Ali et al. (2025)</i>	Narrative review	AI, ML, NLP, digital health tools	AI may support diagnosis, monitoring, personalization and care delivery, but implementation requires attention to privacy, bias, explainability, and governance.	Broad narrative scope; limited direct evidence on long-term clinical outcomes; need for stronger governance frameworks.
Ethical and regulatory safety	<i>WHO (2024)</i>	Policy and guidance document	AI and LLM	WHO emphasizes risks related to safety, privacy, misinformation bias, accountability, and the use of AI among vulnerable populations.	Guidance document rather than empirical study; recommendations require adaptation to local healthcare and regulatory contexts.
Overall evidence base	<i>Rezaei et al. (2026)</i>	Review	AI chatbots, ML, digital phenotyping Emotional AI	Supports screening, psychoeducation, symptom monitoring, emotional support, and personalized interventions.	Insufficient evidence for severe mental illness, chronic disorders, long-term safety, and culturally diverse populations.

Source: Author's compilation based on the studies included in this review. Abbreviations: AI, artificial intelligence; CBT, cognitive behavioral therapy; LLM, large language model; RCT, randomized controlled trial; ML, machine learning; NLP, natural language processing

As shown in Table 2, the available evidence suggests potential benefits of AI-based mental health interventions, particularly in psychoeducation, symptom monitoring, emotional support, and cognitive behavioral therapy-based activities. However, important methodological limitations and safety concerns remain, warranting careful interpretation of current findings.

The most frequently cited benefits include:

1. Emotional support

Social support is considered an important protective factor against negative psychological outcomes, including stress, depressive symptoms, and anxiety (Roohafza et al., 2014; Huang et al., 2021). Social chatbots may provide users with a sense of being understood, emotional validation, empathy, and companionship, even when users are aware of the non-human nature of the interaction (Li et al., 2024). This phenomenon can be interpreted in light of the CASA (Computers Are Social Actors) paradigm, according to which people tend to respond to computers and interactive systems in ways that resemble social responses to other humans (Nass & Moon, 2000).

Generative chatbots can adapt their responses to the user's language, tone, and emotional context, which may enhance engagement, perceived understanding, and a therapeutic-alliance-like relationship. In this context, companion chatbots may help reduce perceived loneliness (Zhang et al., 2025). Chatbots may also be particularly acceptable to digitally native users, including adolescents and young adults. In addition, they may provide low-threshold support for individuals who avoid traditional mental health services because of stigma, fear of diagnostic labeling, or reluctance to seek face-to-face help. People experiencing greater psychological distress and those who avoid labeling or stigmatization may therefore be more inclined to use chatbots as an initial source of support (Rezaei et al., 2026).

2. Psychoeducation, Screening, and CBT-Based Support

Previous research suggests that ChatGPT may demonstrate promising performance in tasks related to the assessment of various mental disorders (Wei et al., 2023; Hadar-Shoval et al., 2023; Sezgin et al., 2023; Wu et al., 2023). For example, Franco D'Souza et al. (2023) analyzed ChatGPT responses to 100 psychiatric clinical cases. In 61 cases, the model received an "A" rating, and the authors reported no misdiagnoses or unacceptable responses. These findings suggest that ChatGPT may demonstrate substantial psychiatric knowledge and the ability to interpret clinical case reports. Similar findings were reported by Schubert et al. (2023), who evaluated the performance of ChatGPT-4.0 using neurological examination-style questions. The model answered 85% of the questions correctly, achieving a higher score than the human average of 73.8%. In another study on the prognosis and long-term consequences of depression, GPT-4, Claude, and Bard demonstrated high agreement with mental health professionals. In the analyzed cases, all models recommended combining psychotherapy with antidepressant pharmacotherapy (Blease et al., 2024).

Rezaei et al. (2026) discuss broader applications of AI beyond chatbots alone. The authors describe AI and machine learning models that analyze text, speech, physiological data, and social media posts to detect linguistic markers of depression or anxiety and to support suicide risk prediction. Such models may assist in symptom assessment, early detection, and risk monitoring (Rezaei et al., 2026). In addition, the authors refer to the concept of "digital phenotyping," which involves the analysis of digital data, such as physical activity, social media activity, smartwatch data, and medical records, to assess mental well-being. For example, reduced physical activity detected by wearable devices may serve as a potential indicator of depressive symptoms (Thakkar et al., 2024). Another relevant area is Emotional AI, which involves the analysis of emotions based on text, voice, facial expressions, gestures, and physiological signals, as well as the delivery of relaxation, breathing, and mindfulness exercises (Thakkar et al., 2024).

AI systems may analyze speech, text, digital behavior, and device-generated data to detect early signals of depression, anxiety, PTSD, and suicide risk. Data from smartphones, social media platforms, and wearable devices may reveal psychological patterns earlier than traditional clinical assessment. For example, AI can analyze tone of voice, speech rate, pauses, word choice, sentence structure, and the emotional tone of written text. Such features may help detect subtle signs of depression, anxiety, and PTSD. Continuous monitoring of steps, sleep, heart rate variability, screen time, geolocation, and social activity may also be used as indicators of mood changes or mental health deterioration. In remote consultations, AI may support clinicians by analyzing a patient's voice, facial expression, and language during a video visit to help identify possible signs of anxiety, depression, or mood deterioration. Ali et al. (2025) also mention chatbots such as Woebot and Wysa as tools that provide coping strategies, mood monitoring, and CBT-based components. (Ali et al., 2025)

AI may also support the personalization of mental health care by tailoring interactions to the user, analyzing text data, mood patterns, and behavioral indicators, and generating more individualized recommendations. Rezaei et al. (2026) discuss natural language processing, personalization, and tailored support as features that may increase user engagement and enhance the therapeutic potential of chatbot-based interventions. AI can also support educational campaigns, analyze social media discourse, and provide users with personalized information about mental health. Thakkar et al. (2024) identify natural language processing, sentiment analysis, and data mining as methods for monitoring public attitudes toward mental health. The authors also indicate that chatbots may conduct empathetic conversations, provide coping strategies, refer users to professional help, and help bridge information gaps related to mental health issues.

Some AI-based systems are grounded in established psychotherapeutic approaches, particularly cognitive behavioral therapy (CBT). These systems aim to support users in identifying negative thought patterns, developing coping strategies, and managing emotional distress through CBT-based techniques (Fitzpatrick et al., 2017). One of the most widely studied examples is Woebot, an AI-based conversational agent based on the principles of cognitive behavioral therapy. In a randomized controlled trial, users who interacted with Woebot for two weeks demonstrated a statistically significant reduction in depressive symptoms compared with users who received psychoeducational materials only (Fitzpatrick et al., 2017). However, these effects were observed in the short term and in a controlled trial setting.

Another example is Wysa, an AI-based conversational platform that combines automated dialogue with evidence-based therapeutic techniques, including cognitive behavioral therapy, mindfulness, and stress management strategies (Inkster et al., 2018). These examples illustrate the potential of chatbot-based interventions as accessible and scalable mental health support tools.

Zhang et al. (2025) also highlight the potential of chatbots to support between-session tasks, including CBT-based and positive psychology exercises such as gratitude practice and self-reflection. In CBT-based interactions, a chatbot may help users recognize automatic thoughts, label emotions, and reframe negative beliefs. Similar conclusions were presented by Thakkar et al. (2024), who suggested that AI may support CBT exercises, skill consolidation between sessions, and progress monitoring. Overall, these findings suggest that AI tools may deliver cognitive-behavioral exercises and support progress tracking, although they should be used as adjunctive tools rather than substitutes for professional psychotherapy.

3. 24/7 Availability and Support in Settings With Limited Access to Specialists

A significant benefit of using chatbots in mental health care is their availability outside specialists' working hours. This may reduce barriers to accessing basic emotional support, particularly for individuals who are unable or reluctant to seek traditional forms of psychological or psychiatric care.

Digital tools based on smart devices and AI-enabled solutions may increase the availability of mental health support because they allow users to access help regardless of location or time of day (Zhang et al., 2025; Zafar et al., 2024). These tools may provide basic emotional support, coping strategies, and exercises inspired by cognitive behavioral therapy (Zafar et al., 2024; Rezaei et al., 2026). Chatbots using natural language processing and conversational artificial intelligence may also enable more personalized interactions with users. This may be particularly important for individuals who have limited access to traditional forms of psychological or psychiatric care (Abd-Alrazaq et al., 2019). Such solutions may also lower the threshold for entering the health care system, as some people may find talking to a chatbot less stressful than seeking direct professional help (WHO, 2022).

From a broader perspective, these technologies may partially reduce the gap in access to mental health care, especially in regions where the number of mental health professionals is insufficient (Abd-Alrazaq et al., 2020; Zhang et al., 2025; Rezaei et al., 2026). However, it should be emphasized that AI chatbots should complement, rather than replace, professional diagnosis and treatment. Studies by Ma et al. (2024) and Heston et al. (2023) suggest that conversational agents supported by large language models may play a supportive role in mental health care, primarily by providing fast, accessible, and non-judgmental support.

Such solutions may be particularly useful for people who, for organizational or social reasons, experience difficulty contacting a therapist, for example because of lack of time, place of residence, professional obligations, or limited availability of specialists. These tools may also be relevant for groups at risk of marginalization, including older adults experiencing loneliness and limited social relationships (Ma et al., 2024; Alessa et al., 2023).

A qualitative analysis conducted by Ma et al. (2024), based on Reddit user responses, indicated that exposure to LLM-based systems may foster greater openness among users and strengthen their self-confidence

through personalized and responsive interactions. Positive findings were also reported in the evaluation of the VHope system, based on the DialoGPT model, whose responses were rated by experts as accurate (67%), natural (78%), and empathetic (79%) (Beredo et al., 2022). Additionally, one study found that ChatGPT achieved higher empathy scores than physicians, which may suggest the acceptability of these tools among users (He et al., 2024).

4. Improved Treatment Outcomes

Interactions with AI chatbots may foster a therapeutic-alliance-like relationship, which is considered an important factor in effective psychotherapy (Flückiger et al., 2012; Martin et al., 2000). However, this does not mean that interaction with a chatbot can be equated with a professional therapeutic relationship. Owing to their ability to respond to the user's language, tone, and emotional context, AI chatbots can create conversations that appear more natural, personal, and responsive. Such communication may increase user engagement (Kumar et al., 2025), strengthen the feeling of being heard and understood (Yin et al., 2024), and support the development of a perceived supportive relationship (Siddals et al., 2024). The importance of this phenomenon can be interpreted in light of the common factors model, according to which the effectiveness of psychotherapy depends not only on specific therapeutic techniques, but also on the quality of the relationship, empathy, perceived safety, and the engagement of the person receiving support (Elvins et al., 2008; Ferrario et al., 2024). From this perspective, social chatbots that generate responses perceived as empathetic, personalized, and emotionally appropriate may facilitate users' disclosure of difficulties and increase trust in the interaction. This may contribute to the reduction of some negative psychological symptoms, although the relationship with AI should be considered only a form of complementary support and not a substitute for professional therapy (Cai et al., 2024).

The available evidence suggests that LLM-based conversational tools may support existing forms of mental health intervention, particularly as adjunctive tools that increase access to basic support and information. However, they should be considered complementary to professional care rather than replacements for it. Chatbots should be integrated into regular health care pathways instead of functioning as standalone substitutes. Zhang et al. (2025) emphasize that future research should examine how generative AI chatbots can be incorporated into existing treatment processes. Generative AI chatbots appear to have the greatest potential as low-threshold supportive tools for psychoeducation, emotional support, self-management, and between-session therapeutic exercises. Nevertheless, current evidence supports their use primarily as adjunctive tools rather than replacements for professional mental health care (Zhang et al., 2025).

Rezaei et al. (2026) reached similar conclusions, emphasizing that AI should complement humans rather than replace mental health specialists. The authors argue that AI systems may create new connections between patients, chatbots, specialists, and digital infrastructure, but these systems must be designed to complement, rather than replace, the therapeutic relationship.

Risks and limitations

While AI-based chatbots may improve the accessibility and scalability of mental health support, their clinical use must be considered in light of substantial safety, ethical, and regulatory concerns (WHO, 2024; Thakkar et al., 2024). The same features that make large language models (LLMs) attractive, including personalization, conversational fluency, and constant availability, may also create risks when these systems are used by vulnerable individuals without adequate safeguards (Zhang et al., 2025; Rezaei et al., 2026).

Despite promising findings on the use of large language models (LLMs) and AI-based chatbots in mental health care, their implementation in clinical practice requires particular caution. One of the main limitations of LLMs is the high dependence of generated responses on prompt wording and contextual information. Even minor differences in the formulation of a question may lead to different, and sometimes inconsistent, recommendations (Spallek et al., 2023; Farhat et al., 2024; Wei et al., 2023; Yongsatianchot et al., 2023; Grabb et al., 2023). Farhat et al. (2024) showed that ChatGPT's treatment-related responses varied depending on how clinical control questions were formulated. Although the model often provided useful responses and usually advised users to consult a specialist, its outputs were not free from errors or potentially inappropriate suggestions. Similar concerns were reported by Perlis et al. (2024), who found that GPT-4 Turbo occasionally recommended pharmacotherapy that experts judged to be less accurate or contraindicated. The literature also describes examples of particularly problematic responses, including suggestions that a person with depression should engage in risky activities to increase pleasure (Grabb et al., 2023).

Several categories of risk are especially relevant in the mental health context. First, generative AI systems may produce unpredictable, inaccurate, or potentially harmful responses (WHO, 2024; Zhang et al., 2025). This is particularly concerning when users are experiencing psychological distress, suicidal ideation, psychosis, trauma-related symptoms, or other severe mental health difficulties (Zhang et al., 2025; Rezaei et al., 2026). In such situations, inappropriate reassurance, false certainty, or inadequate crisis guidance may increase the risk of harm (WHO, 2024; Zhang et al., 2025). Therefore, LLM-based systems should not be used as standalone tools for psychiatric diagnosis, suicide risk assessment, crisis intervention, or emergency psychiatric care (WHO, 2024).

Second, the use of AI chatbots in mental health care raises substantial concerns regarding privacy, confidentiality, and data security (WHO, 2024; Rezaei et al., 2026; Ali et al., 2025). Mental health data are highly sensitive and may include information about symptoms, trauma, medication use, suicidal thoughts, interpersonal relationships, and behavioral patterns collected through applications or wearable devices (WHO, 2024; Ali et al., 2025). Potential risks include inappropriate data storage, unauthorized disclosure, cyberattacks, data breaches, and secondary commercial use of personal information (Rezaei et al., 2026; Ali et al., 2025). Breaches of mental health data may lead to stigma, discrimination, emotional harm, legal consequences, or loss of trust in digital health services (Ali et al., 2025; WHO, 2024). For this reason, AI-based mental health tools should comply with data protection regulations, apply privacy-preserving design, ensure transparent informed consent, and undergo regular security audits (Rezaei et al., 2026; Thakkar et al., 2024).

Third, algorithmic bias and limited cultural generalizability remain important concerns (WHO, 2024; Rezaei et al., 2026; Ali et al., 2025). Many AI models are trained primarily on data from English-speaking, Western, or high-income populations (WHO, 2024; Rezaei et al., 2026). As a result, their performance may be weaker in other linguistic, cultural, and socioeconomic contexts (WHO, 2024; Ali et al., 2025). This may lead to inaccurate interpretation of symptoms, culturally inappropriate recommendations, or the reinforcement of existing health inequalities (Rezaei et al., 2026; Ali et al., 2025). Mental health symptoms and emotional expression vary across cultures, and AI systems that are not adapted to local contexts may fail to recognize clinically relevant distress or may misinterpret culturally specific forms of communication (WHO, 2024). Therefore, culturally sensitive design, local validation, multilingual testing, and the inclusion of diverse populations in model development are essential (WHO, 2024; Rezaei et al., 2026).

Fourth, the limited interpretability of many AI systems creates clinical and ethical challenges (Ali et al., 2025; Thakkar et al., 2024). The so-called “black-box” nature of some models makes it difficult for clinicians and users to understand how a recommendation was generated (Ali et al., 2025). This lack of explainability may reduce clinicians’ trust, limit accountability, and make it harder to identify errors, bias, or unsafe outputs (Ali et al., 2025; Thakkar et al., 2024). In mental health care, where clinical decisions often require contextual understanding, empathy, and nuanced judgment, interpretability is particularly important (Rezaei et al., 2026; Thakkar et al., 2024).

Fifth, there is still insufficient long-term evidence regarding the safety and effectiveness of AI-based mental health tools (Rezaei et al., 2026; Thakkar et al., 2024). Many existing studies are short-term and focus mainly on outcomes measured over several weeks or months (Rezaei et al., 2026). Evidence remains limited regarding long-term effects, relapse prevention, chronic mental disorders, severe psychiatric conditions, and real-world implementation (Rezaei et al., 2026; Ali et al., 2025). AI tools may be useful as adjunctive support for mild or moderate symptoms, psychoeducation, self-monitoring, or between-session engagement, but their effectiveness in complex conditions such as schizophrenia, bipolar disorder, severe depression, substance use disorders, or acute mental health crises remains uncertain (Rezaei et al., 2026; WHO, 2024).

Sixth, AI chatbots may affect the therapeutic relationship and users’ emotional attachment to technology (Zhang et al., 2025; Thakkar et al., 2024). Systems that simulate empathy, warmth, and understanding may increase engagement, but they may also promote excessive anthropomorphization, emotional dependence, or unrealistic expectations about the chatbot’s abilities (Zhang et al., 2025). This is especially relevant for socially isolated users, children, adolescents, and people experiencing mental health crises (Zhang et al., 2025; WHO, 2024). Chatbots designed as virtual companions may reduce loneliness in some users, but they may also increase social withdrawal or replace help-seeking from human professionals if appropriate safeguards are not in place (Zhang et al., 2025).

Seventh, responsibility and accountability remain unresolved (WHO, 2024; Rezaei et al., 2026; Thakkar et al., 2024). If an AI system provides harmful advice, misses a serious symptom, gives an incorrect recommendation, or fails to respond adequately to a crisis, it may be unclear who is responsible: the developer, the healthcare institution, the clinician, the platform provider, or the user (Rezaei et al., 2026; Thakkar et al.,

2024). This accountability gap is particularly problematic in mental health care, where inappropriate recommendations may have serious clinical consequences (WHO, 2024; Thakkar et al., 2024). Clear regulatory frameworks, defined professional responsibilities, transparent documentation, and human supervision are therefore necessary (WHO, 2024; Rezaei et al., 2026).

Finally, AI-based mental health tools must be integrated carefully into clinical workflows (Ali et al., 2025; Thakkar et al., 2024). Even a technically advanced system may be ineffective or unsafe if it does not fit the practice of psychiatrists, psychologists, psychotherapists, or primary care providers (Ali et al., 2025). Implementation should involve mental health professionals, patients, ethicists, and policymakers (WHO, 2024; Thakkar et al., 2024). AI tools should be designed to support, rather than replace, human care and should remain aligned with core therapeutic values, including empathy, respect, patient autonomy, safety, confidentiality, and a patient-centered approach (WHO, 2024; Rezaei et al., 2026; Thakkar et al., 2024).

Overall, current evidence suggests that AI chatbots and LLM-based systems may offer valuable support in mental health care, particularly in psychoeducation, self-monitoring, and improving access to low-intensity support (Thakkar et al., 2024). However, their use is associated with important clinical, ethical, legal, and societal risks (WHO, 2024; Zhang et al., 2025; Rezaei et al., 2026). These technologies should therefore be treated as adjunctive tools requiring validation, regulation, transparency, human oversight, and continuous monitoring rather than as independent providers of mental health care (WHO, 2024; Thakkar et al., 2024).

Ethical challenges

The rapid development of artificial intelligence in healthcare creates new ethical, legal, and governance challenges. These challenges are particularly important in mental health care, where AI-based applications often process highly sensitive personal data and may be used by individuals experiencing psychological distress or mental health crises. Therefore, robust governance frameworks are needed to ensure safety, transparency, accountability, auditability, and the protection of patients' rights (WHO, 2021; European Union, 2024).

In the European Union, the Artificial Intelligence Act (Regulation (EU) 2024/1689) establishes a risk-based regulatory framework for AI systems. Depending on their intended purpose and risk classification, AI systems used in healthcare may be subject to requirements related to risk management, data governance, transparency, human oversight, accuracy, robustness, cybersecurity, and post-market monitoring (European Union, 2024). These requirements are especially relevant for AI tools that may influence clinical decision-making or affect patients' access to care.

International ethical guidance also emphasizes the need for responsible AI governance in healthcare. The World Health Organization identifies six core ethical principles for the use of AI in health: protecting human autonomy; promoting human well-being, human safety, and the public interest; ensuring transparency, explainability, and intelligibility; fostering responsibility and accountability; ensuring inclusiveness and equity; and promoting responsive and sustainable AI (WHO, 2021). In the context of mental health chatbots, these principles mean that users should be clearly informed when they are interacting with an AI system rather than a human professional, should retain control over their data and decisions, and should be protected from unsafe or misleading outputs.

Data protection is one of the central ethical challenges in AI-based mental health care. Conversations with chatbots may include information about depressive symptoms, anxiety, trauma, loneliness, medication use, self-harm, suicidal thoughts, interpersonal difficulties, and behavioral patterns collected through mobile applications or wearable devices. Such information is highly sensitive and may expose users to stigma, discrimination, emotional harm, or other negative consequences if confidentiality is breached (WHO, 2021; Ali et al., 2025). For this reason, privacy-preserving approaches such as anonymization, secure storage, encryption, federated learning, differential privacy, and fairness audits have been proposed as important safeguards for AI systems in mental health care (Ali et al., 2025).

Data protection regulations also play a key role. The General Data Protection Regulation (GDPR) establishes requirements for the lawful processing of personal data, data minimization, transparency, and individual rights, including access, rectification, and erasure. Mental health data constitute data concerning health and therefore fall within a special category of personal data under Article 9 of the GDPR, requiring additional safeguards and specific legal grounds for processing (European Union, 2016; Floridi et al., 2018).

Algorithmic bias is another major ethical concern. AI systems trained primarily on data from English-speaking, Western, or high-income populations may perform less accurately in other linguistic, cultural, and

socioeconomic groups. In mental health care, this is particularly problematic because the expression of distress, emotional symptoms, help-seeking behavior, and communication styles may vary across cultures. As a result, biased or poorly validated systems may misinterpret symptoms, provide culturally inappropriate recommendations, or reinforce existing inequalities in access to care (WHO, 2021; Ali et al., 2025).

Transparency and explainability are also essential. Users and clinicians should understand the limitations of AI systems, including the possibility of factual errors, hallucinations, false reassurance, and inconsistent responses. Generative AI systems and large language models (LLMs) can produce fluent and persuasive responses even when the information is inaccurate or misleading, raising concerns about patient safety, reliability, and clinical decision-making (Hua et al., 2025; Blease et al., 2024). In mental health settings, this risk is particularly important because users may place excessive trust in systems that simulate empathy, understanding, or therapeutic competence.

The use of AI chatbots also raises concerns about emotional dependence, anthropomorphization, and changes in human relationships. Chatbots and virtual companions may provide users with a sense of availability, emotional support, and being understood, but they may also encourage excessive attachment or reduce help-seeking from human professionals. These concerns are especially relevant for socially isolated individuals, adolescents, and people experiencing mental health crises. Reported cases and media discussions have raised concern about harmful human–AI interactions, although causal relationships require careful empirical evaluation and should not be overstated (Bhuiyan, 2025; Duffy, 2024). From an ethical perspective, these risks are closely linked to user autonomy, informed consent, transparency, and the duty to protect vulnerable individuals from foreseeable harm.

Accountability remains a further unresolved issue. If an AI chatbot provides harmful advice, fails to recognize a crisis, gives inaccurate information, or contributes to a delay in professional care, it may be unclear who is responsible: the developer, platform provider, healthcare institution, clinician, or user. Clear accountability structures are therefore necessary before such systems are widely implemented in mental health care (WHO, 2021; WHO, 2024; European Union, 2024).

At the international level, the OECD Principles on Artificial Intelligence emphasize inclusive growth, human-centered values, transparency, robustness, security, safety, and accountability in the development and use of AI systems (OECD, 2023). These principles are consistent with WHO guidance and support the view that AI-based mental health tools should be designed and implemented in a way that protects human autonomy, promotes equity, and strengthens rather than replaces professional care.

Overall, the ethical implementation of AI chatbots in mental health care requires more than technical performance. It requires robust data protection, transparency, bias mitigation, clinical validation, human oversight, professional accountability, and continuous monitoring in real-world settings. Effective governance will likely require collaboration between policymakers, clinicians, researchers, patient organizations, ethicists, and technology developers. Regulatory frameworks should protect patients and promote trust while avoiding unnecessary restrictions on responsible innovation. Achieving this balance is crucial for the safe and ethical integration of AI into mental health care.

Critical appraisal of the current evidence base

Despite the rapid growth of research on artificial intelligence in psychiatric care, the current evidence base remains limited in several important respects. A major concern is the restricted generalisability of many AI models. Numerous studies rely on relatively small, homogeneous, retrospective, or geographically limited datasets, which may reduce model performance when applied to broader and more diverse clinical populations (Shatte et al., 2019). Models trained predominantly on data from high-income countries or specific demographic groups may not perform equally well across different cultural, linguistic, and socioeconomic contexts, raising concerns about algorithmic bias and equitable access to AI-supported mental healthcare.

Methodological limitations are also common. Many studies lack rigorous external validation, use short follow-up periods, or evaluate performance primarily under controlled experimental conditions. This may lead to overestimation of predictive accuracy, particularly when high-dimensional data are analysed using relatively small samples (Rajkomar et al., 2019). In addition, many studies focus on technical outcomes, such as accuracy, sensitivity, specificity, or user engagement, rather than clinically meaningful endpoints. Evidence remains more limited regarding whether AI-based tools improve symptom remission, functional recovery, quality of life, hospitalisation rates, long-term safety, or treatment continuity.

The evidence for AI-based chatbots and digital therapeutic tools is similarly mixed. Although some randomised controlled trials have shown short-term reductions in symptoms of depression and anxiety, many studies include self-selected samples, brief observation periods, and participants with mild to moderate symptom severity (Fitzpatrick et al., 2017; Abd-Alrazaq et al., 2020). Evidence remains particularly limited for adolescents, older adults, culturally diverse populations, individuals with severe mental illness, complex comorbidities, or acute psychiatric crises. Overall, current findings should therefore be interpreted with caution. While AI technologies show considerable promise, stronger prospective, multicentre, externally validated, and clinically oriented studies are needed before widespread implementation in psychiatric care can be recommended.

Future directions and clinical integration

Future development of AI in psychiatric care should prioritise clinical validation, patient safety, transparency, and integration with existing models of care. Large-scale longitudinal studies are needed to assess not only diagnostic or predictive accuracy, but also effectiveness, cost-effectiveness, safety, acceptability, and impact on real-world patient outcomes. Future research should also include more diverse populations, particularly adolescents, older adults, people from different cultural and linguistic backgrounds, and patients with psychiatric conditions beyond depression and anxiety.

Responsible clinical integration will require transparent and explainable AI systems, especially in mental healthcare, where decisions may directly affect patient safety, trust, and therapeutic relationships. AI should be understood as a supportive tool rather than a replacement for clinicians. The most appropriate model is likely to be a hybrid approach, in which AI assists with screening, monitoring, risk detection, and decision support, while clinicians retain responsibility for diagnosis, treatment planning, crisis management, and therapeutic care (Topol, 2019; Torous et al., 2021).

Clear governance frameworks are also necessary. Future implementation should include human oversight, data protection procedures, transparent reporting of model development and validation, regular bias audits, and predefined crisis protocols for situations involving suicide risk, self-harm, or clinical deterioration. Interdisciplinary collaboration between clinicians, data scientists, ethicists, policymakers, and patients will be essential to ensure that AI tools are clinically useful, ethically acceptable, and safe for deployment across different healthcare settings.

Conclusions

Artificial intelligence-based chatbots and large language model systems have considerable potential to support mental health care by improving access to low-intensity support, psychoeducation, self-management, symptom monitoring, and between-session engagement. The evidence reviewed in this article suggests that these tools may be particularly useful in settings where access to mental health professionals is limited, and for individuals who experience barriers related to stigma, geography, cost, or availability of services. AI systems may also support patients by assisting with monitoring, and the personalization of care.

However, the current evidence does not support the use of AI chatbots or LLM-based systems as independent providers of mental health care. Their implementation is associated with significant clinical, ethical, legal, and social risks, including inaccurate or harmful responses, data privacy concerns, algorithmic bias, limited transparency, unclear accountability, emotional dependence, and insufficient validation in diverse clinical populations. These risks are especially important in mental health contexts, where users may be vulnerable, distressed, or at risk of self-harm.

Therefore, AI-based mental health tools should be regarded as adjunctive technologies rather than replacements for psychiatrists, psychologists, psychotherapists, or other mental health professionals. Their safe and ethical use requires robust clinical validation, transparent reporting, privacy-preserving design, bias assessment, human oversight, clear crisis protocols, and compliance with relevant regulatory frameworks. Interdisciplinary collaboration between participants, researchers, ethicists, policymakers, patients, and technology developers will be essential to ensure that these tools are safe, equitable, and clinically meaningful.

In conclusion, AI chatbots and LLM-based systems may play an important supportive role in the future of mental health care, particularly in improving accessibility and scalability. Nevertheless, their integration into clinical practice should proceed cautiously, with continuous monitoring and strong governance mechanisms. Future research should focus on long-term effectiveness, safety, real-world implementation, and outcomes in diverse populations and severe mental health conditions.

Author's contribution:**Conceptualization:** Monika Łapot, Krzysztof Hober, Dominika Krupnik**Methodology:** Monika Łapot, Krzysztof Hober, Joanna Górak, Wojciech Neumann**Software:** Wiktoria Bzdak, Wiktoria Skowron**Validation:** Monika Łapot, Joanna Górak, Wojciech Neumann**Formal analysis:** Monika Łapot, Krzysztof Hober, Mateusz Skowron**Investigation:** Monika Łapot, Krzysztof Hober, Dominika Krupnik, Mateusz Skowron, Wiktoria Bzdak, Patrycja Szewińska**Resources:** Mateusz Skowron, Wiktoria Bzdak, Patrycja Szewińska**Data curation:** Monika Łapot, Dominika Krupnik, Wiktoria Skowron, Mateusz Skowron, Wiktoria Bzdak, Patrycja Szewińska**Writing – rough preparation:** Monika Łapot, Krzysztof Hober, Joanna Górak**Writing – review and editing:** Monika Łapot, Dominika Krupnik, Wojciech Neumann**Visualization:** Monika Łapot, Wiktoria Skowron, Mateusz Skowron, Patrycja Szewińska**Supervision:** Monika Łapot, Wiktoria Bzdak**Project administration:** Monika Łapot, Joanna Górak

All authors have read and agreed with published version of the manuscript.

Funding Statement: This study did not receive special funding.**Conflict of Interest Statement:** The authors declare no conflicts of interest.**Declaration of generative AI and AI-assisted technologies in the writing process:** In preparing this work, the authors used the ChatGPT tool, which helped them translate and improve the language and increase readability. After using this tool, the authors have reviewed and edited the content as needed and take full responsibility for the substantive content of the publication.**REFERENCES**

1. Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, 132, Article 103978. <https://doi.org/10.1016/j.ijmedinf.2019.103978>
2. Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and Meta-Analysis. *Journal of Medical Internet Research*, 22(7), e16021. <https://doi.org/10.2196/16021>
3. Alessa, A., & Al-Khalifa, H. S. (2023). Towards designing a ChatGPT conversational companion for elderly people. In *Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments* (pp. 667–674). Association for Computing Machinery. <https://doi.org/10.1145/3594806.3596572>
4. Ali, M., Ali, S., Abbas, Q., Abbas, Z., & Lee, S. W. (2025). Artificial intelligence for mental health: A narrative review of applications, challenges, and future directions in digital health. *DIGITAL HEALTH*, 11, Article 20552076251395548. <https://doi.org/10.1177/20552076251395548>
5. Arias, D., Saxena, S., & Verguet, S. (2022). Quantifying the global burden of mental disorders and their economic value. *EClinicalMedicine*, 54, 101675. <https://doi.org/10.1016/j.eclinm.2022.101675>
6. Berrezueta-Guzman, S., Kandil, M., Martín-Ruiz, M.-L., Pau de la Cruz, I., & Krusche, S. (2024). Future of ADHD care: Evaluating the efficacy of ChatGPT in therapy enhancement. *Healthcare*, 12(6), Article 683. <https://doi.org/10.3390/healthcare12060683>
7. Bhatt, S. (2025). Digital mental health: Role of artificial intelligence in psychotherapy. *Annals of Neurosciences*, 32(2), 117–127. <https://doi.org/10.1177/09727531231221612>
8. Bhuiyan, J. (2025, August 29). *ChatGPT encouraged adam raine's suicidal thoughts*. His family's lawyer says OpenAI knew it was broken. The Guardian. <https://www.theguardian.com/us-news/2025/aug/29/chatgpt-suicide-openai-sam-altman-adam-raine>
9. Blease, C. R., Locher, C., Gaab, J., Hägglund, M., & Mandl, K. D. (2024). Generative artificial intelligence in primary care: An online survey of UK general practitioners. *BMJ Health & Care Informatics*, 31(1), e101102. <https://doi.org/10.1136/bmjhci-2024-101102>
10. Blease, C., Worthen, A., & Torous, J. (2024). Psychiatrists' experiences and opinions of generative artificial intelligence in mental healthcare: An online mixed methods survey. *Psychiatry Research*, 333, Article 115724. <https://doi.org/10.1016/j.psychres.2024.115724>
11. Bond, R. R., Mulvanna, M. D., Potts, C., O'Neill, S., Ennis, E., & Torous, J. (2023). Digital transformation of mental health services. *npj Mental Health Research*, 2(1), 13. <https://doi.org/10.1038/s44184-023-00033-y>
12. Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, 18(sup1), 37–49. <https://doi.org/10.1080/17434440.2021.2013200>

13. Burdzik, M. (2023). Social stigma of people with mental disorders and attitude to psychiatric treatment in Polish society. *European Psychiatry*, 66(Suppl 1), S463. <https://doi.org/10.1192/j.eurpsy.2023.993>
14. Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3), 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
15. Cai, N., Gao, S., & Yan, J. (2024). How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure. *Humanities and Social Sciences Communications*, 11, Article 687. <https://doi.org/10.1057/s41599-024-03212-0>
16. Centers for Disease Control and Prevention. (2025, June 9). *About mental health*. <https://www.cdc.gov/mental-health/about/index.html>
17. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., & Mitchell, M. (2015). CLPsych 2015 shared task: Depression and PTSD on Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 31–39). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W15-1204>
18. Cruz-Gonzalez, P., He, A. W., Lam, E. P., Ng, I. M. C., Li, M. W., Hou, R., Chan, J. N., Sahni, Y., Vinas Guasch, N., Miller, T., Lau, B. W., & Sánchez Vidaña, D. I. (2025). Artificial intelligence in mental health care: A systematic review of diagnosis, monitoring, and intervention applications. *Psychological Medicine*, 55, e18. <https://doi.org/10.1017/S0033291724003295>
19. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, 71, 10–49. <https://doi.org/10.1016/j.specom.2015.03.004>
20. De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), 128–137. <https://doi.org/10.1609/icwsm.v7i1.14432>
21. Dehbozorgi, R., Zangeneh, S., Khooshab, E., Nia, D. H., Hanif, H. R., Samian, P., Yousefi, M., Hashemi, F. H., Vakili, M., Jamalimoghadam, N., & Lohrasebi, F. (2025). The application of artificial intelligence in the field of mental health: A systematic review. *BMC Psychiatry*, 25(1), 132. <https://doi.org/10.1186/s12888-025-06483-2>
22. Duffy, C. (2024, October 30). 'There are no guardrails.' this mom believes an AI chatbot is responsible for her son's suicide. CNN. <https://edition.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit>
23. Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, 28(7), 1167–1187. <https://doi.org/10.1016/j.cpr.2008.04.002>
24. Elyoseph, Z., Levkovich, I., & Shinan-Altman, S. (2024). Assessing prognosis in depression: Comparing perspectives of AI models, mental health professionals and the general public. *Family Medicine and Community Health*, 12(Suppl 1), Article e002583. <https://doi.org/10.1136/fmch-2023-002583>
25. European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation)*. Official Journal of the European Union, L 119, 1–88. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
26. European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (EC) no 300/2008, (EU) no 167/2013, (EU) no 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (artificial intelligence act)*. Official Journal of the European Union, L, 2024/1689. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
27. Eurostat. (2020, May 6). *Number of psychiatrists: How do countries compare?*. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200506-1>
28. Eurostat. (2025). *Healthcare personnel statistics – physicians*. Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Healthcare_personnel_statistics_-_physicians
29. Farhat, F. (2024). ChatGPT as a complementary mental health resource: A boon or a bane. *Annals of Biomedical Engineering*, 52(5), 1111–1114. <https://doi.org/10.1007/s10439-023-03326-7>
30. Ferrario, A., Sedlakova, J., & Trachsel, M. (2024). The role of humanization and robustness of large language models in conversational artificial intelligence for individuals with depression: A critical analysis. *JMIR Mental Health*, 11, e56569. <https://doi.org/10.2196/56569>
31. Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). Can smartphone mental health interventions reduce symptoms of anxiety? a meta-analysis of randomized controlled trials. *Journal of Affective Disorders*, 218, 15–22. <https://doi.org/10.1016/j.jad.2017.04.046>
32. 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), e19. <https://doi.org/10.2196/mental.7785>
33. Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
34. Flückiger, C., Del Re, A. C., Wampold, B. E., Symonds, D., & Horvath, A. O. (2012). How central is the alliance in psychotherapy? a multilevel longitudinal meta-analysis. *Journal of Counseling Psychology*, 59(1), 10–17. <https://doi.org/10.1037/a0025749>

35. Franco D'Souza, R., Amanullah, S., Mathew, M., & Surapaneni, K. M. (2023). Appraising the performance of ChatGPT in psychiatry using 100 clinical case vignettes. *Asian Journal of Psychiatry*, 89, Article 103770. <https://doi.org/10.1016/j.ajp.2023.103770>
36. Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: Mixed-Method systematic review. *JMIR Mental Health*, 6(10), e14166. <https://doi.org/10.2196/14166>
37. GBD 2023 Mental Disorder Collaborators (2026). Updated trends in the global prevalence and burden of mental disorders, 1990-2023: A systematic analysis for the global burden of disease study 2023. *Lancet (London, England)*, 407(10543), 2040–2064. [https://doi.org/10.1016/S0140-6736\(26\)00519-2](https://doi.org/10.1016/S0140-6736(26)00519-2)
38. Grabb, D. (2023). The impact of prompt engineering in large language model performance: A psychiatric example. *Journal of Medical Artificial Intelligence*, 6, Article 20. <https://doi.org/10.21037/jmai-23-71>
39. Habicht, J., Dina, L.-M., McFadyen, J., Stylianou, M., Harper, R., Hauser, T. U., & Rollwage, M. (2025). Generative AI-enabled therapy support tool for improved clinical outcomes and patient engagement in group therapy: Real-world observational study. *Journal of Medical Internet Research*, 27, Article e60435. <https://doi.org/10.2196/60435>
40. Hadar-Shoval, D., Elyoseph, Z., & Lvovsky, M. (2023). The plasticity of ChatGPT's mentalizing abilities: Personalization for personality structures. *Frontiers in Psychiatry*, 14, Article 1234397. <https://doi.org/10.3389/fpsy.2023.1234397>
41. He, W., Zhang, W., Jin, Y., Zhou, Q., Zhang, H., & Xia, Q. (2024). Physician versus large language model chatbot responses to web-based questions from autistic patients in chinese: Cross-sectional comparative analysis. *Journal of Medical Internet Research*, 26, Article e54706. <https://doi.org/10.2196/54706>
42. He, Y., Yang, L., Qian, C., Li, T., Su, Z., Zhang, Q., & Hou, X. (2023). Conversational agent interventions for mental health problems: Systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research*, 25, e43862. <https://doi.org/10.2196/43862>
43. Henderson, C., Evans-Lacko, S., & Thornicroft, G. (2013). Mental illness stigma, help seeking, and public health programs. *American Journal of Public Health*, 103(5), 777–780. <https://doi.org/10.2105/AJPH.2012.301056>
44. Heston, T. F. (2023). Safety of large language models in addressing depression. *Cureus*, 15(12), Article e50729. <https://doi.org/10.7759/cureus.50729>
45. Hua, Y., Liu, F., Yang, K., Li, Z., Na, H., Sheu, Y.-H., Zhou, P., Moran, L. V., Ananiadou, S., Clifton, D. A., Beam, A., & Torous, J. (2025). Large language models in mental health care: A scoping review. *Current Treatment Options in Psychiatry*, 12, Article 27. <https://doi.org/10.1007/s40501-025-00363-y>
46. 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 mHealth and uHealth*, 6(11), e12106. <https://doi.org/10.2196/12106>
47. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
48. Kumar, A., Shankar, A., Behl, A., Chakraborty, D., & Gundala, R. R. (2025). Anthropomorphic generative AI chatbots for enhancing customer engagement, experience and recommendation. *Journal of Consumer Marketing*, 42(4), 472–483. <https://doi.org/10.1108/JCM-06-2024-6922>
49. Li, H., & Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts of close encounters with AI chatbots. *Journal of Computer-Mediated Communication*, 29(5), Article zmae015. <https://doi.org/10.1093/jcmc/zmae015>
50. Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *npj Digital Medicine*, 6(1), 236. <https://doi.org/10.1038/s41746-023-00979-5>
51. Liu, A. R., Pataranutaporn, P., & Maes, P. (2024). *Chatbot companionship: A mixed-methods study of companion chatbot usage patterns and their relationship to loneliness in active users*. arXiv. <https://doi.org/10.48550/arXiv.2410.21596>
52. Ma, Z., Mei, Y., & Su, Z. (2023). Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. *AMIA Annual Symposium Proceedings*, 2023, 1105–1114.
53. Maples-Keller, J. L., Bunnell, B. E., Kim, S.-J., & Rothbaum, B. O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry*, 25(3), 103–113. <https://doi.org/10.1097/HRP.0000000000000138>
54. Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68(3), 438–450.
55. Minister Zdrowia. (2021). *Obwieszczenie ministra zdrowia z dnia 27 sierpnia 2021 r. w sprawie mapy potrzeb zdrowotnych*. Dziennik Urzędowy Ministra Zdrowia, 2021(69). https://dziennikmz.mz.gov.pl/eli/DUM_MZ/2021/69/ogl
56. Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences*, 25(2), 113–122. <https://doi.org/10.1017/S2045796015001067>
57. Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
58. OECD & European Observatory on Health Systems and Policies. (2023). *Poland: Country health profile 2023*. State of Health in the EU. OECD Publishing. <https://doi.org/10.1787/f597c810-en>
59. Oh, H., Pickering, T. A., Schiffman, J., Soffer-Dudek, N., LaBrie, J. W., & Pedersen, E. R. (2024). Web-based personalized normative feedback to decrease stigma and increase intentions to seek help. *Stigma and Health*, 10(3), 420–427. <https://doi.org/10.1037/sah0000518>

60. O'Keefe, M., White, K., & Jennings, J. C. (2021). Asynchronous telepsychiatry: A systematic review. *Journal of Telemedicine and Telecare*, 27(3), 137–145. <https://doi.org/10.1177/1357633X19867189>
61. Organisation for Economic Co-operation and Development (OECD). (2021). *A new benchmark for mental health systems: Tackling the social and economic costs of mental ill-health*. OECD Health Policy Studies. OECD Publishing. <https://doi.org/10.1787/4ed890f6-en>
62. Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M. J. D., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., ... Unützer, J. (2018). The Lancet commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553–1598. [https://doi.org/10.1016/S0140-6736\(18\)31612-X](https://doi.org/10.1016/S0140-6736(18)31612-X)
63. Patel, V., Xiao, S., Chen, H., Hanna, F., Jotheeswaran, A. T., Luo, D., Parikh, R., Sharma, E., Usmani, S., Yu, Y., Druss, B. G., & Saxena, S. (2016). The magnitude of and health system responses to the mental health treatment gap in adults in india and china. *The Lancet*, 388(10063), 3074–3084. [https://doi.org/10.1016/S0140-6736\(16\)00160-4](https://doi.org/10.1016/S0140-6736(16)00160-4)
64. Perlis, R. H., Goldberg, J. F., Ostacher, M. J., & Schneck, C. D. (2024). Clinical decision support for bipolar depression using large language models. *Neuropsychopharmacology*, 49(9), 1412–1416. <https://doi.org/10.1038/s41386-024-01841-2>
65. Prince, M., Patel, V., Saxena, S., Maj, M., Maseko, J., Phillips, M. R., & Rahman, A. (2007). No health without mental health. *Lancet (London, England)*, 370(9590), 859–877. [https://doi.org/10.1016/S0140-6736\(07\)61238-0](https://doi.org/10.1016/S0140-6736(07)61238-0)
66. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *The New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMr1814259>
67. Rezaei, Z., Khorraminia, A., Shi, D., & Banad, Y. M. (2026). Network-based artificial intelligence in mental healthcare: A systematic review of chatbots, artificial intelligence/machine learning models and ethical considerations in global healthcare networks. *DIGITAL HEALTH*, 12, Article 20552076261421688. <https://doi.org/10.1177/20552076261421688>
68. Roohafza, H. R., Afshar, H., Keshteli, A. H., Mohammadi, N., Feizi, A., Taslimi, M., & Adibi, P. (2014). What's the role of perceived social support and coping styles in depression and anxiety?. *Journal of Research in Medical Sciences: the Official Journal of Isfahan University of Medical Sciences*, 19(10), 944–949.
69. Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare (Basel, Switzerland)*, 11(6), 887. <https://doi.org/10.3390/healthcare11060887>
70. Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., ... Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, 398(10312), 1700–1712. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7)
71. Schubert, M. C., Wick, W., & Venkataramani, V. (2023). Performance of large language models on a neurology board-style examination. *JAMA Network Open*, 6(12), Article e2346721. <https://doi.org/10.1001/jamanetworkopen.2023.46721>
72. Schwartz, E. (2024, May 19). *The global mental health crisis: 10 numbers to note*. Project HOPE. <https://www.projecthope.org/news-stories/story/the-global-mental-health-crisis-10-numbers-to-note/>
73. Sezgin, E., Chekeni, F., Lee, J., & Keim, S. (2023). Clinical accuracy of large language models and Google search responses to postpartum depression questions: Cross-sectional study. *Journal of Medical Internet Research*, 25, Article e49240. <https://doi.org/10.2196/49240>
74. Sharma, G., Yaffe, M. J., Ghadiri, P., Gandhi, R., Pinkham, L., Gore, G., & Abbasgholizadeh-Rahimi, S. (2025). Use of artificial intelligence in adolescents' mental health care: Systematic scoping review of current applications and future directions. *JMIR Mental Health*, 12, e70438. <https://doi.org/10.2196/70438>
75. Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, 49(9), 1426–1448. <https://doi.org/10.1017/S0033291719000151>
76. Siddals, S., Torous, J., & Coxon, A. (2024). “It happened to be the perfect thing”: Experiences of generative AI chatbots for mental health. *npj Mental Health Research*, 3, Article 48. <https://doi.org/10.1038/s44184-024-00097-4>
77. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., ... Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, 620(7972), 172–180. <https://doi.org/10.1038/s41586-023-06291-2>
78. Spallek, S., Birrell, L., Kershaw, S., Devine, E. K., & Thornton, L. (2023). Can we use ChatGPT for mental health and substance use education? examining its quality and potential harms. *JMIR Medical Education*, 9, Article e51243. <https://doi.org/10.2196/51243>
79. Thakkar, A., Gupta, A., & De Sousa, A. (2024). Artificial intelligence in positive mental health: A narrative review. *Frontiers in Digital Health*, 6, Article 1280235. <https://doi.org/10.3389/fgdh.2024.1280235>
80. Thornicroft, G. (2006). *Shunned: Discrimination against people with mental illness*. Oxford University Press.
81. Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.
82. Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 20(3), 318–335. <https://doi.org/10.1002/wps.20883>

83. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, 64(7), 456–464. <https://doi.org/10.1177/0706743719828977>
84. Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet. Psychiatry*, 3(2), 171–178. [https://doi.org/10.1016/S2215-0366\(15\)00505-2](https://doi.org/10.1016/S2215-0366(15)00505-2)
85. Wei, Y., Guo, L., Lian, C., & Chen, J. (2023). ChatGPT: Opportunities, risks and priorities for psychiatry. *Asian Journal of Psychiatry*, 90, Article 103808. <https://doi.org/10.1016/j.ajp.2023.103808>
86. World Health Organization. (2021). *Mental health atlas 2020*. World Health Organization. <https://www.who.int/publications/i/item/9789240036703>
87. World Health Organization. (2022). *World mental health report: Transforming mental health for all*. World Health Organization. <https://www.who.int/publications/i/item/9789240049338>
88. World Health Organization. (2023, May 16). *WHO calls for safe and ethical AI for health*. <https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health>
89. World Health Organization. (2025). *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. World Health Organization. <https://www.who.int/publications/i/item/9789240084759>
90. Wu, Y., Chen, J., Mao, K., & Zhang, Y. (2023). Automatic post-traumatic stress disorder diagnosis via clinical transcripts: A novel text augmentation with large language models. In *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (pp. 1–5). IEEE. <https://doi.org/10.1109/BioCAS58349.2023.10388714>
91. Yin, Y., Jia, N., & Waksalak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences of the United States of America*, 121(14), e2319112121. <https://doi.org/10.1073/pnas.2319112121>
92. Yongsatianchot, N., Ghanad Torshizi, P., & Marsella, S. (2023). Investigating large language models' perception of emotion using appraisal theory. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 1–8). IEEE. <https://doi.org/10.1109/ACIIW59127.2023.10388194>
93. Zafar, F., Fakhare Alam, L., Vivas, R. R., Wang, J., Whei, S. J., Mehmood, S., Sadeghzadegan, A., Lakkimsetti, M., & Nazir, Z. (2024). The role of artificial intelligence in identifying depression and anxiety: A comprehensive literature review. *Cureus*, 16(3), e56472. <https://doi.org/10.7759/cureus.56472>
94. Zhang, Q., Zhang, R., Xiong, Y., Sui, Y., Tong, C., & Lin, F.-H. (2025). Generative AI mental health chatbots as therapeutic tools: Systematic review and meta-analysis of their role in reducing mental health issues. *Journal of Medical Internet Research*, 27, Article e78238. <https://doi.org/10.2196/78238>
95. Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. *Journal of Affective Disorders*, 356, 459–469. <https://doi.org/10.1016/j.jad.2024.04.057>